

DOCUMENT RESUME

ED 043 182

EC 030 165

AUTHOR Robinson, James T.; Tolman, Richard R.
TITLE A Formative Evaluation of Me Now, Unit I, Digestion and Circulation, Life Sciences for the Educable Mentally Handicapped Intermediate Grades (11-13 years).
INSTITUTION Colorado Univ., Boulder.
SPONS AGENCY Office of Education (DHEW), Washington, D.C.
PUB DATE Sep 70
NOTE 106p.
EDRS PRICE EDRS Price MF-\$0.50 HC-\$5.40
DESCRIPTORS Biological Sciences, Course Evaluation, Curriculum Evaluation, *Educable Mentally Handicapped, Evaluation Methods, *Exceptional Child Research, Experimental Curriculum, *Science Education, *Textbook Evaluation
IDENTIFIERS Biological Sciences Curriculum Study

ABSTRACT

In order to evaluate Unit I (Digestion and Circulation) of the Me Now series of Life Sciences for the Educable Mentally Handicapped, 139 students in an experimental group (mean IQ 72.04; mean CA 144.9 months) and 154 control subjects (mean IQ 70.2; mean CA 148.3 months) were pre- and post-tested following instruction in the Me Now series (experimental group) or regular curriculum (controls). Results showed that age, WISC total IQ, race and cerebral dysfunction are not significant predictors of success on the posttest. The pretest, however, was a significant (P less than .01) predictor of success, as was the sex factor (P less than .05, males scoring higher). The data indicated that two-thirds of the items focused on baseline information and that prior knowledge of the concepts was the best determinant of high scores on the posttest. (RD)

A Formative Evaluation of ME NOW

Unit I, Digestion and Circulation

ED0 43182

Life Sciences for the Educable Mentally Handicapped
Intermediate Grades (11-13 years)



experimental materials prepared by the

BIOLOGICAL SCIENCES CURRICULUM STUDY
University of Colorado



September 1970

EC 030 165 E

ED0 43182

EC030165

A Formative Evaluation of Me Now, Unit I, Digestion and Circulation,
Life Sciences for the Educable Mentally Handicapped
Intermediate Grades (11-13 years)

Experimental Materials

Prepared by

the

Biological Sciences Curriculum Study

University of Colorado

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE
PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS
STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION
POSITION OR POLICY.

September 1970

EC08010

0810001

The project presented or reported herein was performed pursuant to a grant from the U. S. Office of Education, Department of Health, Education, and Welfare. The opinions expressed herein, however, do not necessarily reflect the position or policy of the U. S. Office of Education, and no official endorsement by the U. S. Office of Education should be inferred.

LIFE SCIENCES FOR THE EDUCABLE MENTALLY HANDICAPPED

Advisory Committee

WILLIAM R. REID, Chairman
University of Florida
Gainesville, Florida

ROBERT L. ERDMAN
University of Utah
Salt Lake City, Utah

WARREN HEISS
Yeshiva University
New York, New York

JERRY H. HUBSCHMAN
Wright State University
Dayton, Ohio

GEORGE SHEPERD
University of Oregon
Eugene, Oregon

Experimental Group Teachers

JOE BEATTY
Foothills Elementary School
Boulder, Colorado

MARY LOU DE GUIRE
Foster Elementary School
Arvada, Colorado

ROGER O. DRAKE
Baker Elementary School
Denver, Colorado

HELEN FULLER
Milliken Middle School
Milliken, Colorado

RUTH GADDIS
Miller Special School
Lakewood, Colorado

FLORENCE HAAG
Condon Elementary School
Eugene, Oregon

NANCY JOHNSON
Greenlee Elementary School
Denver, Colorado

NIKKI JONES
Miller Special School
Lakewood, Colorado

JOAN SEBASTIAN
Jackson Elementary School
Salt Lake City, Utah

BETTY TOMER
Putnam Elementary School
Fort Collins, Colorado

CARNEICE WHITE
Ebert Elementary School
Denver, Colorado

Evaluation Committee

OLIVER P. KOLSTOE, Chairman
Colorado State College
Greeley, Colorado

JEAN BEARD
San Jose State College
San Jose, California

JOE M. STEELE
University of Illinois
Urbana, Illinois

Test Writing Committee

JEAN BEARD
San Jose State College
San Jose, California

ANNE CARROLL
University of Denver
Denver, Colorado

JACK CAWLEY
University of Connecticut
Storrs, Connecticut

AUSTIN J. CONOLLY
University of Missouri
Columbia, Missouri

ROY GROMME
Nicolet High School
Milwaukee, Wisconsin

CARY REICHART
University of Florida
Gainesville, Florida

HILL WALKER
University of Oregon
Eugene, Oregon

Control Group Teachers

FRANCIS BARTELL
Ocala Junior High School
Ocala, Florida

BARBARA GUERRIERO
Daniels Elementary School
Malden, Massachusetts

ETHEL HUDSON
Roper Elementary School
San Jose, California

DAVID LEVAY
Washington Elementary School
Colorado Springs, Colorado

JOHN MEHLE
Rose Hill Elementary School
Commerce City, Colorado

MARCIA MINTER
Rose Hill Elementary School
Commerce City, Colorado

PATRICIA PROCIOT
Fort King Elementary School
Ocala, Florida

RUTH ROGERS
Lyndale Elementary School
San Jose, California

SYRIL RUSHAY
Taylor Elementary School
Colorado Springs, Colorado

MRS. SHEEHAN
Daniels Elementary School
Malden, Massachusetts

MARY THOMAS
Lyndale Elementary School
San Jose, California

BSCS Staff

WILLIAM V. MAYER
Director

MANERT H. KENNEDY
Associate Director

GEORGE M. CLARK
Assistant Director

KEITH BUMSTED
Business Manager

PATRICK E. BALCH
Consultant

THOMAS J. CLEAVER
Consultant

ROY O. GROMME
Consultant

FRED A. RASMUSSEN
Consultant

JAMES T. ROBINSON
Consultant

HAROLD A. RUPERT
Consultant

RICHARD R. TOLMAN
Consultant

ROBERT F. WILSON
Art Director

LAWRENCE STRAND
Artist

ROY UDO
Artist

PREFACE

This report presents data and interpretations of data that are relevant to classroom trials of Me Now, Unit I, Digestion and Circulation, the first set of materials prepared in the Life Sciences for the Educable Mentally Handicapped Program of the Biological Sciences Curriculum Study.

We wish to thank members of the BSCS Staff and the Evaluation Committee for their most helpful suggestions in the preparation of this report. We especially wish to acknowledge the assistance of Dr. J. Steele, University of Illinois, for his insightful assistance throughout the project.

The completeness of the data reflects the dedication of the experimental and control group teachers to the project. The interpretations of the data are the sole responsibility of the project evaluators.

James T. Robinson

Richard R. Tolman

CONTENTS

LIFE SCIENCES FOR THE EDUCABLE MENTALLY HANDICAPPED.	<i>i</i>
PREFACE	<i>ii</i>
LIST OF TABLES	<i>iv</i>
LIST OF FIGURES	<i>vi</i>

CHAPTER	Page
I. A LIFE SCIENCES PROGRAM FOR THE EDUCABLE MENTALLY HANDICAPPED	1
The Instructional Program	
Design of the Formative Evaluation	
Data Gathering Instruments and Procedures	
II. DATA ANALYSIS AND RESULTS	13
Students	
Teacher Reports	
Objective Achievement Tests: Descriptive Data and Interpretations	
Objective Achievement Tests: Multiple Stepwise Regression Analysis, Experimental Group	
Objective Achievement Tests: Analysis of Covariance, Experimental Group	
Objective Achievement Tests: Residual Gain Analyses Experimental Group	
Objective Achievement Tests: Analysis of Variance	
Objective Achievement Tests: Items and Objectives	
Objective Achievement Tests: Cluster Analysis	
Objective Achievement Tests: Discussion	
III. SUMMARY AND RECOMMENDATIONS	77
Were the program objectives achieved?	
What were the results of the formative evaluation?	
What is recommended?	
REFERENCES CITED	97

LIST OF TABLES

TABLE	Page
1. Formative Evaluation Model	5
2. Data Collection Program	9
3. Experimental Group, Student Background Information, April, 1970	14
4. Control Group, Student Background Information, April, 1970	16
5. Experimental Group, Item Analysis of Pretest and Posttest Forms A and B	18
6. Frequency Distributions of Weighted Raw Scores, Forms A and B, Experimental and Control Groups	20
7. Experimental Group Pretest and Posttest Data.	21
8. Control Group Pretest and Posttest Data	22
9. Results of Multiple Linear Regression Analysis, Experimental Group	24
10. Matrix of Correlation Coefficients, Experimental Group	26
11. Multiple Stepwise Regression Analysis, Experimental Group	27
12. Analysis of Covariance Between Classes on Adjusted Posttest Means, Pretest as Covariate	28
13. N, Means, Standard Deviations, and Adjusted Means, Experimental Group	29
14. Residual Gain and Class Data, Experimental Group	31
15. ANOVA Experimental Group, Residual Gain Blocked on Verbal IQ	32
16. Mean Residual Gain Scores and Standard Deviations, Experimental Group, Blocked on Verbal IQ	32

LIST OF TABLES--continued

TABLE	Page
17. ANOVA Experimental Group, Residual Gain Blocked on Performance IQ	33
18. Mean Residual Gain Scores and Standard Deviations, Experimental Group, Blocked on Performance IQ	33
19. ANOVA, Experimental Group	34
20. Male and Female Mean Residual Gain Scores and Standard Deviations, Experimental Group	34
21. Posttest Means and Standard Deviations, Experimental and Control Groups	36
22. ANOVA, Experimental and Control Posttest Scores	36
23. Item Responses and Biserial Correlations for Experimental and Control Groups, Forms A and B	39
24. Pretest to Posttest Changes, Item Pair A-1, B-1	42
25. Pretest to Posttest Changes, Item Pair A-6, B-7	43
26. Pretest to Posttest Changes, Item Pair A-8, B-8	44
27. Pretest to Posttest Changes, Item Pair A-4, B-14	45
28. Pretest to Posttest Changes, Item Pair A-13, B-4	47
29. Pretest to Posttest Changes, Item Pair A-3, B-15	48
30. Pretest to Posttest Changes, Item A-10	49
31. Pretest to Posttest Changes, Item Pair A-20, B-10	50
32. Pretest to Posttest Changes, Item Pair A-12, B-27	52
33. Pretest to Posttest Changes, Item Pair A-5, B-9	54
34. Pretest to Posttest Changes, Item Pair A-9, B-17	56
35. Pretest to Posttest Changes, Item Pair A-7, B-6	57
36. Pretest to Posttest Changes, Item Pair A-18, B-24	59
37. Pretest to Posttest Changes, Item Pair A-19, B-22	60

LIST OF TABLES--continued

TABLE	Page
38. Pretest to Posttest Changes, Item B-13	62
39. Pretest to Posttest Changes, Item Pair A-15, B-25	63
40. Pretest to Posttest Changes, Item Pair A-26, B-26	64
41. Summary of Item Objectives, Level, and Gain for Cluster 1, Posttest B	67
42. Summary of Item Objectives, Level, and Gain for Cluster 2, Posttest B	68
43. Correlation of Items with the Oblique Cluster Domains, Posttest A	69
44. Summary of Item Objectives, Level, and Gain for Cluster 1, Posttest A	71
45. Correlation of Items with the Oblique Cluster Domains, Posttest A	72

LIST OF FIGURES

Figure 1	66
Figure 2	70

CHAPTER I

A LIFE SCIENCES PROGRAM FOR THE EDUCABLE MENTALLY HANDICAPPED

The Instructional Program

The development of materials for educable mentally handicapped youngsters is a new curriculum development venture for the Biological Sciences Curriculum Study. The project was initiated in the summer of 1969 with the selection of an Advisory Committee and the assembling of staff consultants, secretarial staff, etc. Recommendations were made by the Advisory Committee that the first curriculum component be devised for the intermediate grades, 11 to 13 year-old group, and deal with the human body. Materials were to be addressed directly to the designated age group and its state of development.

Advisory Committee members and the BSCS staff consultants proposed that a field test of a small part of the program be conducted in the spring of 1970 if at all possible, rather than to develop a larger part of the program without evaluation. This procedure was adopted, materials were produced and field trials were conducted in eleven classrooms from April to June, 1970.

The program of curriculum development for the educable mentally handicapped is based on the assumptions that:

1. most teachers of the Educable Mentally Handicapped are not science teachers and need specific direction in the use of inquiry strategies in the teaching of science concepts.
2. E.M.H. children need and can respond effectively to an activity-centered instruction approach.

3. to achieve the objectives, materials should provide a balance between detail and motivation; that is, the amount of minute and abstract detail that can be learned is probably a function of the interest and motivation that can be established to deal with it.
4. the classroom environment and the materials should be uncluttered with distractors; however, a variety of perceptual modes and instructional media should be used in all efforts at communication, e.g., sight, touch, smell, etc.
5. an activity must involve the student in ways of applying the desired behavior; transfer cannot be assumed.
6. activities should be developed in small discrete units that build on or reinforce a concept or skill.
7. entry points should be concerned with concrete, tangible "things" rather than with abstract, intangible ideas or concepts.
8. ideas must be developed without the necessity for reading on the part of the student.
9. vocabulary, where possible, should involve functional language rather than technical nomenclature.
10. learning for the E.M.H. student requires slower pacing, greater redundancy and time for participation by each student.
11. efforts to describe the "average" E.M.H. child are essentially futile because of variability within the population, therefore, materials and methods should allow for attention to individual differences and needs.¹

Additional information on the intent of the materials is stated in "The Involvement of Inquiry Objectives in the Instructional Program". This inquiry model, as described below, is assumed to be an effective model when adapted to the particular characteristics of children in intermediate grade (age 11-13) E.M.H. classes.

¹Biological Sciences Curriculum Study. Me Now, Unit I, Digestion and Circulation (Experimental Edition). Boulder, Colorado: The Regents of the University of Colorado, Spring 1970, p.iii (Hereafter referred to as Me Now.)

Scientific knowledge is dynamic, growing and changing as man develops ways of viewing phenomena and asks new questions about himself and his environment. This dynamic quality of science should and can be reflected in this program of instruction in several ways.

Throughout the program, students should confront real phenomena in such ways that their curiosity is aroused and their further investigation is encouraged. Thereby, students become involved in inquiry behaviors such as observing, describing, inferring, associating, and applying. The student who is actively involved in investigation to satisfy his own curiosity develops knowledge that is his own.

For E.M.H. students, no less than for others, the excitement of discovery adds meaning and value to learning. It is not always feasible, however, to allow each student to rediscover knowledge for himself. When investigation by the student is not feasible, the information necessary for him to make interpretations must be provided in an objective manner. The student should be given the opportunity to distinguish between facts and ideas, names and conceptions; he should be made aware of what are simplifications of more complex structures.

The application of inquiry methodology and techniques to the learning activities of the E.M.H. student is, at this point, largely based upon intuition. The program must, of course, be tested in the classroom to determine its effect. Until the results of this test are in, we urge that teachers and others involved in this testing project avoid prejudging the extent or limits of the E.M.H. student's capabilities. The instructional program has been designed in sequences that are increasingly complex; only experience can tell us what the upper limits of complexity are for E.M.H. students.

Two hierarchies are presumed to exist in these materials. One pertains to level of cognitive difficulty and the other to the sequencing of content elements. The objectives may range across the four levels of cognitive difficulty. In most instances, it is assumed that the E.M.H. child will gain from a design that proceeds from simple to complex, from concrete to abstract, from what is familiar to what is unfamiliar. The design of activities for this unit of studies proceeds on that assumption. However, depending on the nature of the objective, a child may, for example, be expected to associate objectives that are familiar to him without first describing them.

PRESUMED LEVELS OF COGNITIVE DIFFICULTY

4.	APPLICATIONAL	PREDICTING	HIGHEST
3.	RELATIONAL	INFERRING ASSOCIATING	
2.	CONCEPTUAL	COMPARING IDENTIFYING	
1.	PERCEPTUAL	DESCRIBING OBSERVING	LOWEST

With regard to the hierarchy of sequence, it is assumed that the child should be allowed to study single elements of a system before being asked to describe or work with interactions of those elements. Similarly, a child might be expected to associate a name with those functions. Staff judgment is being initially applied to the sequencing of content in the hope that guidelines may develop from analysis of the results.²

Design of the Formative Evaluation

A formative evaluation, modeled on that proposed by Stake, for initial field testing was prepared as the materials were developed.³ The design of the evaluation study is shown in Table 1.

Eleven trial classes were selected primarily from schools in the immediate vicinity of the BSCS (Boulder, Colorado) office. This was done to achieve economy in teacher expenses and staff expense in assisting classes with instructional problems. Two classes outside the state of Colorado were included to extend the trial group. Administration of the pretests and posttests to eleven control classes both within and outside of Colorado was included in the design. Classes for the control

² Me Now, p.iv.

³ Robert E. Stake. "The Countenance of Educational Evaluation." Teachers College Record 68:523-540 (April), 1967.

TABLE 1
Formative Evaluation Model

Antecedents: conditions existing prior to teaching and learning that may be related to outcomes.

Questions	Intentions	Data to be Collected	Data Collection Source	Instruments to be Developed	Method of Data Processing
1. Did the students in the test classes have the characteristics of the population for whom the materials were developed?	Students in test classes will be between ages 11 and 13 years and will range in total IQ from 50 to 80.	Student age, I.Q., race, cerebral dysfunction, father's occupation.	School records.	Data collection form.	Computer: descriptive statistics (BMD01V)
2. Can the student perform the objective prior to instruction?	Students would respond to most test items in a random fashion on a pretest.	Pretest over objectives.	Parallel forms of a multiple choice test.	Forms A & B of test over objectives.	Computer: Fortap program Descriptive statistics

TABLE 1 (continued)

Transactions: the encounters of student with teacher, student with materials, teacher with materials, and student with student.

Questions	Intentions	Data to be Collected	Data Collection Source	Instruments to be Developed	Method of Data Processing
3. Did the teacher use the strategies as prescribed?	Teachers will use the strategies and add to them when necessary.	Teacher reports Video tape	Teacher Video tape	Report form Protocol	Hand compiling ?
4. Did the teacher modify the strategies?	Teachers will not modify, but will supplement selected strategies.	Teacher reports Video tape	Teacher Video tape	Report form Protocol	Hand compiling ?
5. How much time is devoted to a particular block or unit of material?	Four to six weeks time at 30 minutes per day should be adequate.	Teacher reports Video tape	Teacher Video tape	Report form Protocol	Hand compiling ?
6. Did the intended student behavior occur?	Students will be able to perform the intended behaviors.	Teacher reports Video tape	Teacher Video tape	Pupil evaluation form Protocol	Hand compiling ?
7. Did the intended student behavior occur to an acceptable degree?	Students will want to perform the activities presented in the materials.	Teacher reports Video tape	Teacher Video tape	Pupil evaluation form Protocol	Hand compiling ?
8. Did unintended behaviors occur?	Lack of interest and inability to perform will be limited to only a small number of students in each class.	Teacher reports Video tape	Teacher Video tape	Pupil evaluation form Report form Protocol	Hand compiling Hand compiling ?
9. What was the relationship between the teacher behaviors and student behaviors?	Teacher behaviors as prescribed will elicit the desired student responses.	Teacher reports Video tape	Teacher Video tape	Report form Protocol	Hand compiling ?

TABLE 1 (continued)
Outcomes: the impact of instruction on teachers, students,
administrators and parents.

Questions	Intentions	Data to be Collected	Data Collection Source	Instruments to be Developed	Methods of Data Processing
10. What were the teacher's reactions to the materials?	Teachers will respond critically to each of sixteen sections of the materials.	Teacher reports	Teacher	Report form	Hand compiling
11. What were the teacher's reactions to the teacher strategies?	Teachers will support the inclusion of strategies (with modifications) in the <u>Guide</u> .	Teacher reports	Teacher	Report form	Hand compiling
12. What were the student's reactions to the materials?	Teachers will report on student reactions in terms of attention, interest and performance for each of sixteen sections of the materials.	Teacher reports	Teacher	Pupil evaluation form	Hand compiling
13. What were the student's reactions to the teacher strategies?	Students will perform the instructional objectives and will show gains on posttest scores.	Teacher reports Pretest-posttest scores	Teacher Form A and Form B Pre-test and Posttest	Pupil evaluation form Multiple choice test	Hand compiling Computer: Fortap program ANCOVA, multiple stepwise regression, residual gain, ANOVA

TABLE 1 (continued)
Outcomes (continued): the impact of instruction on teachers, students, administrators and parents.

Questions	Intentions	Data to be Collected	Data Collection Source	Instruments to be Developed	Methods of Data Processing
14. Did the students meet the objectives?	Each student will have the opportunity to perform the instructional objectives; a high percentage will be able to perform as stated.	Pretest-posttest scores	Form A and Form B Pre-test and Posttest	Multiple choice test	Computer: Fortap program ANCOVA, multiple stepwise regression, residual gain, ANOVA Hand Compiling
15. Did the students meet the objectives to a suitable degree?	Each student will attain a posttest score similar to a predicted score.	Pretest-posttest scores	Form A and Form B Pre-test and Posttest	Multiple choice test	Computer: Fortap program ANCOVA, multiple stepwise regression, residual gain, ANOVA

group were to be intermediate grade students using their regular curriculum.

Pretests were administered to both experimental and control classes during the weeks of April 13 and 20, 1970. Instruction with the experimental materials began the day following pretesting in the experimental classes. Experimental teachers were briefed at a two-day session. They were asked to proceed at their own pace, and to administer the posttest on the last possible day of the school year. Control teachers were also asked to administer the posttest on the last possible day of the school year. The data collection model is shown in Table 2.

TABLE 2
Data Collection Program

Group	Pretest	Treatment	Posttest	Other Data
Experimental 11 classes 139 students	6 classes Form A	BSCS <u>Me Now</u> , Unit I	6 classes Form B	Student Characteristics
	5 classes Form B	Sixteen sets of teacher report forms	5 classes Form A	Videotapes of
Control 11 classes 154 students	5 classes Form A	Regular curriculum No reports	5 classes Form B	Student Characteristics
	6 classes Form B		6 classes Form A	

Data Gathering Instruments and Procedures

Background Student Information

Information about students that might influence interest, attention and performance during instruction was desired. Also, data were needed to determine if the students in the experimental and control groups met target population criteria (see Question 1, Table 1). Verbal, performance and total IQ scores on the WISC for each student, secured within one year of April, 1970, were considered to be important data. Other data to be secured for each student were date of birth, sex, race, medical evidence of cerebral dysfunction, and father's occupation.

In a spot check of E.M.H. classes in several school districts, IQ data were found to vary greatly in the time of administration. Many IQ data were several years old. Lack of funds precluded administration of the WISC to students with scores secured prior to April, 1969.

Date of birth, race, and father's occupation were secured by the teacher of each class from school records. Again, especially in the case of the father's occupation, dates of entry indicated that these data may lack current validity.

A response of "yes" to cerebral dysfunction required that the report of a qualified physician certified to the dysfunction and that such certification be found in the school records. A staff-developed form was provided to experimental and control group teachers for reporting student information.

Teacher Forms

Sixteen sets of feedback forms of two types were bound into the materials at specific points in the instructional sequence. Feedback

reports were spaced so that reporting over a very limited segment of the instructional program could be secured. Questions for checking yes-no responses were provided to secure data regarding the instructional materials, the teacher strategies, and the pupil responses. In addition, teachers were asked to rate each student on a three-point scale regarding attention, interest and performance.

The report form also contained space for reporting the number of days used in teaching each segment, the length of instructional periods and the attendance of each student. Standard procedures were adopted for recording withdrawal and entry of students during the experimental period.

Objective Test

Student performance capabilities for the behaviors specified in the experimental materials were established by means of multiple-choice objective test instruments. The initial criteria recommended for such instruments were that reading not be necessary, that two-choice and three-choice items predominate, and that mixed colored paper be used to enable students to keep track of where they should be on the test.

An initial test writing group prepared over 160 items, many with two picture choices with the question to be read orally by the teacher. Many items suggested were verbal analogies, such as, "A is to B as C is to D or E. Mark the correct word." This writing group proposed that four-choice items were too difficult for the target population.

Two parallel tests (Form A and B) of 30 items each were produced for trial as pretests and posttests. The final test forms consisted of items with four choices (with the exception of two, three-choice items), generally pictures, with the question and instructions for marking to be read

by the test administrator. Two-choice items were rejected as being too simple and as having too high a guess rate. Verbal analogies were rejected as being inconsistent with the program objectives.

Two items on each test form required the use of color. The problem of color for these items and the problem of assisting examinees in keeping their places during the test was solved by printing the test booklets in black and white, but by preparing a black and white or colored slide for each test item. Projection of each item was presumed to be a solution to keeping students from losing their places and to make color production limited to only a few slides.

Each experimental group and control group teacher was provided with a Kodak Carousel projector with the slides in test order and with a detailed procedure for administering the test.

Classroom Transactions

The unique and explicit teacher strategies presented in Me Now led to question 3 to 9 in Table 1. Optimum data for answering these questions could be obtained by videotaping and by using trained observers with protocols for gathering reliable observation data. Time and funds precluded the preparation of protocols and the training and support of classroom observers. Arrangements were made to videotape the complete experimental program of two teachers. Videotapes were to be used on an informal basis by the BSCS staff and were to be made available to the materials revision group. Development of protocols for analysis of the videotapes was to be done if staff time permitted.

CHAPTER II

DATA ANALYSIS AND RESULTS

Students

Intermediate grade youngsters in special education classes are generally characterized as being 11 to 13 year olds who have exhibited evidence of learning difficulties in regular school classes and who have scored from 50 to 80 on individually administered intelligence tests. Complete data for the experimental groups are shown in Table 3. Of 139 students in eleven classes who were in the experimental group the week prior to the initiation of instruction, background data, pretest scores and posttest scores were secured for 122.

The experimental group ranged in age from 117 to 174 months with a mean of 144.9 months as shown in Table 3. No students in the experimental group had measured IQ scores⁴ below 50, but 14 had scores above 80. The mean total IQ for the experimental group was 72.04.

Ethnically, 60.4 percent of the experimental group students were whites, 33.1 percent were Spanish-American, and 6.5 percent were blacks.

Three of the eleven experimental classes were composed of students from inner-city ghettos in Denver, Colorado and Salt Lake City, Utah. One class was composed of students from several school districts in an agricultural area of northern Colorado. Three classes were located in the university communities of Boulder and Fort Collins, Colorado and

⁴Stanford-Binet IQ scores were converted to WISC Full Scale IQ scores using Table 25 in John E. Horrocks, Assessment of Behavior. Columbus, Ohio: Charles E. Merrill Books, 1964, p. 210.

TABLE 3
Experimental Group, Student Background Information, April 1970

		Age in Months					IQ (WISC)					Race **		Cerebral Dysfunction:				
Teacher #	N	M	S	Range	Over 13 yrs	Under 11 yrs	σM	M	S	Range	σM	Over 80	Under 50	1	2	3	Yes	No
Form A*																		
01	11	151.55	6.49	138-160	3	0	1.96	73.64	6.43	64-85	1.91	1	0	7	4	0	4	7
02	15	145.80	12.80	126-172	4	2	3.30	68.00	7.23	54-80	1.87	0	0	6	9	0	3	12
05	12	130.08	5.30	122-136	0	5	1.53	73.83	5.97	64-85	1.72	1	0	12	0	0	6	6
08	13	141.00	13.30	120-168	1	3	3.69	76.23	5.54	68-85	1.54	3	0	5	8	0	0	13
10	11	148.55	9.40	131-164	2	1	2.83	71.91	9.94	50-84	3.00	2	0	11	0	0	2	9
11	13	139.54	9.46	117-151	0	2	2.62	69.92	8.21	58-91	2.28	1	0	13	0	0	4	9
All A	75	142.61	11.95	117-172	10	13	1.38	72.09	7.61	50-91	0.88	8	0	54	21	0	19	56
Form B*																		
03	10	140.90	10.34	122-155	0	2	3.27	72.50	7.00	58-84	2.21	1	0	1	4	5	0	10
04	14	148.79	10.27	135-168	3	0	2.74	70.36	7.61	58-87	2.03	1	0	1	11	2	0	14
06	14	151.00	11.52	134-168	5	0	3.08	74.07	8.47	54-88	2.26	2	0	11	3	0	4	10
07	14	147.29	13.46	132-166	6	0	3.06	71.57	9.88	56-82	2.64	2	0	14	0	0	2	12
09	12	148.00	13.03	122-174	2	1	3.76	71.42	5.62	58-78	1.62	0	0	3	7	2	0	12
All B	64	147.56	11.90	122-174	16	3	1.49	71.97	7.82	54-88	0.98	6	0	30	25	9	6	60
All	139	144.89	12.14	117-174	26	16	1.03	72.04	7.68	50-91	0.65	14	0	84	46	9	25	116
Boys	90	143.91	11.98	122-172	15	11	1.26	73.33	7.20	54-91	0.76	11	0	56	28	6	15	76
Girls	49	146.69	12.34	117-174	11	5	1.76	69.65	8.04	50-85	1.15	3	0	28	18	3	10	40

*Teachers are grouped by the pretest form to which they were assigned, i.e., Form A teachers administered Form A as a pretest and Form B as a posttest.

**White = 1, Mexican-American = 2, Black = 3.

Eugene, Oregon. Four classes were from suburban communities surrounding Denver, Colorado.

Of 154 students in eleven classes in the control group, background data, pretest scores and posttest scores were secured for 118 (see Table 4). This group was older than the experimental group with a mean age of 148.3 months and an age range of 116 to 199 months. Ten students had measured IQ scores of over 80, no students had scores below 50, and the mean total IQ was 70.2.

Ethnically, 58.3 percent of the control group students were whites, 30.6 percent were Spanish-American, and 11.1 percent were blacks.

The data gathered support the interpretation that both the experimental and control groups were members of the population for whom the materials were designed. Limited financial support for evaluation did not permit random selection of either experimental or control groups and, therefore, nothing can be stated regarding the representativeness of either group in terms of the national 11 to 13 year old E.M.H. population.

Teacher Reports

Summaries of the reports of the experimental group teachers for each segment of instruction are available to interested persons from the BSCS offices. These reports include the complete comments of teachers on each instructional segment. Summary interpretations of these data are provided in the next section of this report.

TABLE 4

Control Group, Student Background Information, April 1970

Age in Months										IQ (WISC)						Race**			Cerebral Dysfunction	
Teacher #	N	M	S	Range	Over 13 yrs	Under 11 yrs	σ_M	M	S	Range	σ_M	Over 80	Under 50	1	2	3	Yes	No		
Form A																				
52	17	139.72	6.27	126-148	0	3	1.48	69.17	6.11	57-80	1.48	0	0	0-14-3	0	17				
53	16	138.12	7.80	128-152	0	5	1.95	69.56	6.02	57-77	1.50	0	0	3-11-2	0	16				
56	17	145.53	8.40	125-155	0	2	2.04	67.00	7.30	56-77	1.77	0	0	5-8-4	5	12				
57	11	148.91	9.62	129-161	2	1	2.90	68.64	6.96	52-78	2.10	0	0	8-2-1	1	10				
59	13	138.85	13.14	116-157	1	4	2.41	68.23	8.68	53-81	2.41	1	0	13-0-0	1	12				
All A	74	141.92	9.77	116-161	3	15	1.12	68.51	6.87	52-81	0.80	1	0	29-35-10	7	67				
Form B*																				
51	12	177.83	10.66	162-199	12	0	3.08	67.08	8.82	51-81	2.54	1	0	8-0-4	0	12				
54	10	172.50	14.12	152-199	9	0	4.46	73.80	7.16	59-82	2.26	2	0	10-0-0	1	9				
55	14	142.21	6.09	127-151	0	1	1.63	73.43	4.43	67-81	1.18	1	0	10-4-0	1	13				
58	15	148.20	9.50	139-169	2	0	2.45	71.33	10.68	51-91	2.76	2	0	14-0-1	8	7				
60	14	141.21	6.09	133-157	1	0	1.63	72.71	5.59	60-81	1.50	1	0	8-4-2	0	14				
61	15	157.27	5.44	149-166	9	0	1.41	72.60	6.06	60-83	1.56	2	0	11-4-0	0	15				
All B	80	155.11	16.20	127-199	33	1	1.81	71.85	7.51	51-91	0.84	9	0	61-12-7	10	70				
All	154	148.29	14.91	116-199	36	16	1.19	70.25	7.38	51-91	0.60	10	0	90-47-17	17	137				
Boys	84	148.55	14.78	116-199	20	10	1.61	70.81	7.33	52-91	0.80	6	0	46-29-9	11	73				
Girls	70	149.04	15.17	121-199	16	6	1.79	69.57	7.43	51-83	0.89	4	0	44-18-8	6	64				

*Teachers are grouped by the pretest form to which they come assigned, i.e., Form A teachers administered Form A as a pretest and Form B as a posttest.

**White = 1, Mexican-American = 2, Black = 3.

Objective Achievement Tests:

Descriptive Data and Interpretations

Pretests were administered to experimental classes on April 13 and 14, 1970. Slide trays were then retrieved and sent to control class teachers. Teachers were asked to pretest each student and to retain the slide tray until students who were absent could be tested. This delay, plus extremely poor mail deliveries, delayed getting slide trays to control group teachers. The initial plan to complete pretest administration in experimental and control classes during the week of April 13, 1970 could not be achieved. The actual pretest administration pattern was not completed in control classes until the week of April 20.

Additional slide trays and minor corrections of items were prepared for posttesting. Wide differences in school closing dates and the completion of instruction spread out the administration of posttesting from May 18, when one teacher completed the materials, to June 6, when the last posttest was administered. The shortest time between pretest and posttest administration was five weeks in both the experimental and control groups.

Item analyses, using the Fortap program,⁶ were prepared for both experimental and control group pretests. Items were grouped by the objectives they were designed to assess. The Evaluation Committee used these data to recommend items to be discarded and those to be retained

⁵Frank B. Baker and T. J. Martin. Fortap: A Fortran Test Analysis Package, Laboratory of Experimental Design, Wisconsin Research and Development Center for Cognitive Learning. The University of Wisconsin, March 1, 1968.

TABLE 5

Experimental Group, Item Analysis of Pretest and Posttest Forms A and B
Me Now, Unit I. Digestion and Circulation N = 122

Obj #	Item No. A B		Disposition and Comment	Pretest				Posttest			
				Percent Correct A B		Forms A & B		Percent Correct A B		Forms A & B	
						Total Items	%C			Total Items	%C
1	1	1	Retain-Revise to make both items eating	62	48	2	56	74	62	2	67
	2	2	Omit-Not related to the objective								
	6	7	Retain-Revise 6C to a glove, improve 7D	22	43	2	31	44	50	2	48
	Total, Objective 1			42	46	4	44	59	56	4	58
2	8	8	Retain. Baseline.	91	96	2	94	98	96	2	97
3	4	14	Retain-(Drop on next test?) Baseline.	96	96	2	96	94	94	2	94
4	13	4	Retain. Baseline.	75	78	2	76	74	82	2	79
5	3	15	Retain-Revise, smashes not parallel.	65	50	2	58	72	62	2	66
	10	19	Retain 10. Not parallel. Omit 19.	16	-	1	16	61	-	1	61
	16	3	Omit-Cut and bite are not parallel.								
	20	10	Retain-Art work is too striking.	66	80	2	72	91	93	2	92
Total, Objective 5			49	65	5	55	75	78	5	73	
6	12	27	Retain. Baseline.	76	78	2	77	89	87	2	88
	23	20	Omit-Implies single part.								
Total, Objective 6			76	78	2	77	89	87	2	88	
7	5	9	Retain-Item hangs on "just."Reword stem.	38	17	2	29	50	65	2	58
8	9	17	Retain. Relational.	59	56	2	57	80	69	2	74
	17	5	Omit-Information not important.								
Total, Objective 8			59	56	2	57	80	69	2	74	
9	7	6	Retain-Revise 6 Baseline.	63	72	2	67	85	91	2	89
11	11	16	Omit-Negative.Rotate tables, change to positive.								
	18	24	Retain. Baseline.	76	78	2	77	70	81	2	76
	19	22	Retain. Recall	66	50	2	59	72	79	2	76
	24	18	Omit-Revise to make positive.								
	28	11	Omit-Doesn't get at anything.								
Total, Objective 11			71	64	4	68	71	80	4	76	
12	14	21	Omit-Doesn't get at function.								
	22	23	Omit-Unimportant.								
Total, Objective 12											
14	25	13	Retain 13-Item 25 not taught. Recall	-	06	1	06	-	22	1	22
	27	12	Omit-Symbol-object confusion.								
	29	29	Omit-Incomprehensible								
Total, Objective 14			-	06	1	06	-	22	1	22	
16	15	25	Retain. Application.	10	20	2	15	33	43	2	38
	21	30	Omit-Inconsistent with materials.								
	26	26	Retain Application.	35	43	2	38	30	24	2	26
	30	28	Omit-Inconsistent with materials.								
Total, Objective 16			23	32	4	26	32	34	4	33	
(16)	(16)		Total Test	57.2	56.9	32	54.4	69.8	68.8	32	67.7

for further analysis. The disposition of items, arranged by objectives and pretest and posttest item scores for the experimental group are shown in Table 5. Fourteen items on each test form were rejected and omitted from any further analysis. Criteria for rejection were:

1. does the item sample the objective?
2. is the item free from linguistic or visual cues that would render the item useless?
3. does the item sample an important aspect of the instructional program, or does the item provide important baseline data that could be used as diagnostic information?

The data and analyses to be reported are based upon two sixteen-item parallel test forms. Correct response choices for each item were given a weight of 4, omits and incorrect responses were weighted 0, and multiple responses were weighted 1. The maximum possible score for the revised Form A and Form B tests was 64 raw score points each.

Raw score frequency distributions on the pretest and posttest for both experimental and control groups are shown in Table 6. Tables 7 and 8 provide more detailed descriptive data on the pretest and posttest measurements for experimental and control classes. The interpretations that follow are based upon the data provided in these tables.

1. Students using Forms A and B in the experimental classes had the same pretest means. Their unadjusted posttest means were well within the standard error of measurement of the instruments. For sixteen-item tests, the posttest reliabilities were reasonable. Future additions to each test form should assure acceptable test reliabilities, i.e., reliabilities above 0.70.

TABLE 6

Frequency Distributions of Weighted Raw Scores
for
Test Forms A and B, Experimental and Control Groups

Weighted Raw Score	Pretest Groups				Posttest Groups			
	Experimental		Control		Experimental		Control	
	Form A	Form B	Form A	Form B	Form A	Form B	Form A	Form B
64								
60-63				3	5	3	2	
56-59	2	0	1	3	4	4	4	
52-55	1	3	1	8	11	11	6	
48-51	7	3	0	13	9	14	8	8
44-47	7	9	5	11	5	13	12	6
40-43	15	8	12	11	3	6	11	8
36-39	12	10	10	8	9	11	8	12
32-35	8	9	9	2	4	0	3	10
28-31	8	9	8	2	1	2	6	9
24-27	4	1	5	0	1	1	0	1
20-23	4	2	2	1	2	2	2	0
16-19			2			0		1
12-15			1			1		0
8-11								0
4-7								0
0-3								1
Totals	68	54	56	62	54	68	62	56

TABLE 7

Experimental Group Pretest and Posttest Data, June 1970
Unit I, Me Now, Digestion and Circulation

Teacher	Age in Months				Total IQ (WISC)				Pretest						Posttest					
	N	M	S	σ_M	M	S	σ_M	N	M	M%	S	r	σ_M	N	M	M%	S	r	σ_M	
A Form																				
	11	151.55	6.49	1.96	73.64	6.43	1.91	11	41.82	65.34	9.01		2.72	11	49.09	76.70	6.72		2.02	
	01																			
	15	145.80	12.80	3.30	68.00	7.23	1.87	12	31.67	49.48	8.44		2.44	12	40.75	63.67	13.27		3.83	
	02																			
	05	130.08	5.30	1.53	73.83	5.97	1.72	11	39.64	61.94	9.26		2.79	11	43.64	68.19	9.20		2.77	
08	141.00	13.30	3.69	76.23	5.54	1.54	13	34.92	54.56	7.77		2.16	13	42.15	65.86	7.41		2.06		
10	148.55	9.40	2.83	71.91	9.94	3.00	10	36.00	56.25	7.30		2.31	10	45.20	70.63	10.84		3.43		
11	139.54	9.46	2.62	69.92	8.21	2.28	11	39.55	61.80	6.23		1.88	11	44.00	68.75	8.90		2.68		
All A	75	142.61	11.95	1.38	72.09	7.61	0.88	68	37.13	58.00	8.51	0.38	6.47	68	44.01	68.77	9.63	0.59	5.98	
B Form																				
	10	140.90	10.34	3.27	72.50	7.00	2.21	10	38.40	60.00	8.04		2.54	10	44.90	70.16	8.04		2.36	
	03																			
	14	148.79	10.27	2.74	70.36	7.61	2.03	13	33.85	52.89	7.05		1.95	13	45.23	70.67	8.85		2.46	
	04																			
	06	151.00	11.52	3.08	74.07	8.47	2.26	10	42.40	66.25	7.11		2.25	10	52.40	81.88	6.65		2.10	
07	147.29	13.46	3.06	71.57	9.88	2.64	13	37.38	58.41	8.36		2.32	13	42.00	65.63	12.62		3.50		
09	148.00	13.03	3.76	71.42	5.62	1.62	8	35.12	54.88	5.67		2.00	8	42.00	65.63	10.25		3.62		
All B	64	147.56	11.90	1.49	71.97	7.82	0.98	54	37.32	58.10	7.72	0.31	6.22	54	45.24	70.00	9.43	0.61	5.99	
All	139	144.89	12.14	1.03	72.04	7.68	0.65	122	37.21	58.14	8.14		0.74	122	44.56	69.63	9.74		0.88	
Boys	90	143.91	11.98	1.26	73.33	7.20	0.76	78	38.26	59.78	8.57		0.97	78	46.38	72.47	9.25		1.05	
Girls	49	146.69	12.34	1.76	69.65	8.04	1.15	44	35.36	55.25	6.94		1.07	44	41.48	64.81	9.83		1.52	

*Standard errors of estimate (σ_s) for individual scores.

Table 8

Control Group Pretest and Posttest Data, June, 1970
Unit I, Me Now, Digestion and Circulation

Teacher		Age in Months				Total IO (WISC)				Pretest					Posttest				
#	N	M	S	σ_M	M	S	σ_M	N	M	M%	S	r	σ_M	N	M	M%	S	r	σ_M
A																			
Form																			
52	17	139.72	6.27	1.48	69.17	6.11	1.48	15	31.93	49.89	6.78		1.75	15	34.93	54.58	5.55		1.43
53	16	138.12	7.80	1.95	69.56	6.02	1.50	12	36.50	57.03	7.62		2.20	12	40.50	63.28	7.63		2.20
56	17	145.53	8.40	2.04	67.00	7.30	1.77	11	32.00	50.00	9.63		2.90	11	32.36	50.56	12.83		3.87
57	12	148.85	9.60	2.66	68.75	6.65	1.92	7	36.00	56.25	11.31		4.28	7	37.14	58.03	7.90		2.99
59	12	138.83	13.72	3.96	68.08	9.05	2.61	11	33.91	52.98	8.90		2.68	11	35.27	55.11	9.26		2.79
All A	74	142.10	9.77	1.12	68.51	6.87	0.80	56	33.82	52.84	8.51	0.40	6.40	56	35.96	56.19	8.92	0.52	6.12
B																			
Form																			
51	12	177.83	10.66	3.08	67.08	8.82	2.54	9	41.78	65.28	8.51		2.84	9	43.11	73.59	10.73		3.58
54	10	172.50	14.12	4.46	73.80	7.16	2.26	10	40.40	63.13	7.65		2.42	10	41.20	64.38	9.62		3.04
55	14	142.21	6.09	1.63	73.43	4.43	1.18	10	41.60	65.00	5.72		1.81	10	40.00	62.50	6.25		1.98
58	15	148.20	9.50	2.45	71.33	10.68	2.76	12	42.33	66.14	9.87		2.85	12	39.67	61.98	9.72		2.81
60	14	141.21	6.09	1.63	72.71	5.59	1.50	10	34.00	53.13	7.56		2.39	10	38.40	60.00	9.88		3.12
61	15	157.27	5.44	1.41	72.60	6.06	1.56	11	41.09	64.20	5.96		1.80	11	49.45	77.27	4.48		1.35
All B	80	155.11	16.20	1.81	71.85	7.51	0.84	62	40.26	62.91	7.92	0.27	6.56	62	42.00	65.63	9.13	0.40	6.66
All	154	148.78	14.91	1.19	70.25	7.38	0.60	118	37.20	58.13	8.78		0.81	118	39.14	61.16	9.49		0.87
Boys	84	148.55	14.78	1.61	70.81	7.33	0.80	65	37.92	59.25	9.29		1.15	65	40.09	62.64	10.45		1.30
Girls	70	149.04	15.17	1.79	69.57	7.43	0.89	53	36.32	56.75	8.13		1.12	53	37.96	59.31	8.10		1.11

*Standard errors of estimate (σ_s) for individual scores.

2. Students using test Forms A and B in the experimental classes had similar posttest means. A multiple stepwise regression to determine the effects of independent variables was performed. Examination of the multiple regression output to determine the most influential variable(s) followed by an analysis of covariance was performed to determine if there were significant differences between classes on posttest scores, with the pretest score held as a covariate.
3. Control group students using Form B as a pretest performed better than those using Form A. However, the numerical gain in raw scores between pretest and posttest scores was well within the standard error of measurement of the test instruments. The control groups, then, did not change their performance from pretesting to posttesting. Although the above interpretation would lead one to conclude that there would be no significant difference between experimental and control groups, a one-way analysis of variance was computed that confirmed this interpretation.

Objective Achievement Tests:

Multiple Stepwise Regression Analysis,

Experimental Group

In order to determine the effect on posttest scores, if any, of the independent variables, the following question was investigated: "Is there a significant difference in the level of achievement on the reduced (16 items) posttest among students in E.M.H. classes having different

background variables?

The independent variables used to test this question were sex, age, WISC total IQ, evidence of cerebral dysfunction, race, and pretest score. Since Forms A and B are parallel forms, all student scores were pooled and treated as the results of one test.

The test statistic used in testing the question above is the F-statistic generated for each independent variable in the last step of the multiple stepwise regression:

$$F_{\alpha(v_1, v_2)} = \left(\frac{\beta_i}{S\beta_i} \right)^2$$

where: β_i represents the weight of the independent variable (slope of the regression line).

$S\beta_i$ represents the standard error of the weight of the independent variable.

The results for the posttest administered to the 122 students in the experimental group are summarized in Table 9.

TABLE 9
Results of Multiple Linear Regression Analysis,
E.M.H. Experimental Group. N = 122

Independent Variable	β_i	$S\beta_i$	F
Sex	-3.4666	1.6363	4.4887*
Age	0.1075	0.0706	2.3190
WISC Total IQ	0.1815	0.1099	2.7240
Race	1.2423	1.2148	1.0457
Cerebral Dysfunction	1.5089	1.2909	1.3663
Pretest	0.4883	0.1011	23.3321**

*Significant at the .05 level, $F_{.05(1,115)} = 3.93$

**Significant at the .01 level, $F_{.01(1,115)} = 6.87$

The F-value for each independent variable determines the level at which that variable is a significant predictor of a score on the reduced (16-item) posttest instrument.

Discussion

The data indicate that age, WISC total IQ, race and cerebral dysfunction are not significant predictors of success on the reduced posttest. The pretest, however, is a highly significant ($P < .01$) predictor of success on the posttest. Sex is also a significant ($P < .05$) predictor of success on the posttest. These results indicate that prior knowledge of the concepts measured by the test instrument was the best determinant of whether or not the E.M.H. student attained a high score on the posttest. Test analysis shows that 2/3 of the items were aimed at baseline information, hence this conclusion is not surprising, especially since the few achievement items were for the enabling objectives and were at a low cognitive level.

On the IBM cards, male students were coded with a 1 and females with a 2. Since the correlation coefficient between sex and posttest scores was $-.251$ (see Table 10) the investigators inferred that male E.M.H. students scored significantly higher ($P < .05$) than female E.M.H. students on the posttest instrument. An analysis of variance was computed on residual gain scores to confirm this inference.

TABLE 10
Matrix of Correlation Coefficients, E.M.H. Experimental Group

	Age	Total IQ	Race	Cerebral Dysfunction	Pretest	Posttest
Sex	.157	-.230	.049	.002	-.171	-.251
Age		-.227	.026	-.018	.044	.083
Total IQ			-.079	.126	.323	.291
Race				-.002	-.161	-.002
Cerebral Dysfunction					.105	.150
Pretest						.486

The effect of the pretest accounts for approximately 23.6 percent of the variance in the regression equation. This datum is determined by the multiple r^2 in Table 11. The combination of pretest and sex as independent or predictor variables accounts for approximately 26.5 percent of the variance in the regression equation. From Table 11 we find that the inclusion of all six independent variables accounts for only 30.6 percent of the total variance.

TABLE 11
Multiple Stepwise Regression Analysis,
E.M.H. Experimental Group

Step Number	Variable Entered	Multiple r	r^2	Increase in r^2	F-Value to Remove	No. of Independent Variables Included
1	Pretest	.4855	.2357	.2357	37.0164	1
2	Sex	.5144	.2647	.0289	4.6781	2
3	Total IQ	.5264	.2771	.0125	2.0370	3
4	Age	.5399	.2915	.0144	2.3799	4
5	Cerebral Dysfunction	.5478	.3000	.0085	1.4080	5
6	Race	.5535	.3063	.0063	1.0457	6

Since only approximately 30.6 percent of the variance in the regression equation has been accounted for by all of the independent variables, it appears that there should be other important determinants of success on the posttest instrument than those included in this study. This finding may reflect the weakness of the test and indicate that better achievement measures need to be developed. A different interpretation would suggest that the five independent variables secured are inadequate. Different data about the student and characteristics associated with the teacher are two potential sources of variance.

Effects due to student-student and student-teacher interactions are also important independent variables that should be investigated. The proper way to gather these data is either by classroom visitations by someone thoroughly trained in an interaction analysis protocol or by recording the classroom events on video tape and analyzing the tapes at a later date. Unfortunately, this type of analysis is more expensive

than the statistical analysis applied to multiple-choice paper and pencil-type tests. However, in the case of E.M.H. students, where traditionally used independent variables apparently account for a small part of the total variance in the regression equation, the expense of studying classroom transactions in order to determine what actually accounts for the success of these students seems warranted.

Objective Achievement Tests:

Analysis of Covariance, Experimental Group

To investigate the question, "Is there a significant difference in the level of achievement on the reduced (16-item) posttest between E.M.H. classes when adjusted for differences in pretest scores?" an analysis of covariance was performed on students' posttest scores in 11 classes, holding pretest scores as the covariate.

The results of the analysis of covariance, indicating no significant difference between classes on the adjusted posttest means, are summarized in Table 12.

TABLE 12

Analysis of Covariance Between Classes on Adjusted Posttest Means,
Pretest as Covariate

Source	d.f.	Mean Square	F Ratio
Between Groups	10	61.0608	0.8222
Within Groups	110	74.2653	

$F_{.05 (10,110)} = 1.92$ no significant difference

Had there been significant differences among the experimental classes the discovery of contributing factors would have been important. These could have included differences in teacher ability, variation in strategies used, etc. As there were no significant differences, apparently there was little variation as measured by the test. Again more sensitive measures might have revealed differences.

Table 13 summarizes the means and standard deviations for pretests and posttests for each class in the experimental group.

TABLE 13

N, Means, Standard Deviations, and Adjusted Means of 11 Classes,
E.M.H. Experimental Group

Class Number	N	Posttest Mean	Posttest Standard Deviation	Posttest Adjusted Mean	Pretest Mean
1	11	49.09	6.71	46.54	41.82
2	12	40.75	13.27	43.82	31.67
3	10	44.90	7.49	44.24	38.40
4	13	45.23	8.85	47.10	33.85
5	11	43.64	9.20	42.29	39.64
6	10	52.40	6.65	49.53	42.40
7	13	42.00	12.62	41.91	37.38
8	13	42.15	7.41	43.42	34.92
9	8	42.00	10.25	43.16	35.13
10	10	45.20	10.84	45.87	36.00
11	11	44.00	8.90	42.71	39.55

Objective Achievement Tests:

Residual Gain Analyses, Experimental Group

In order to minimize the effect of regression to the posttest mean that is inherent in a pretest-posttest design, a residual gain score was calculated and computer-punched for each student. Residual gain is defined as the difference between a predicted score ($y' = a + bX$) and actual score (Y) on the posttest, or $Y - Y'$, where:

Y' = predicted posttest score

a = Y intercept of the regression line

b = slope of the regression line (the pooled regression coefficient in this case)

X = actual pretest score

Y = actual posttest score.⁶

Table 14 summarizes the description of residual gain data for the 11 classes and 122 students in the E.M.H. experimental group.

⁶Kenneth D. Hopkins. "Regression and the Matching Fallacy in Quasi-Experimental Research," Mimeographed. Boulder, Colorado: Laboratory of Educational Research, University of Colorado, n.d.

TABLE 14
Residual Gain, Class Data, E.M.H. Experimental Group

Class Number	N	Mean	Standard Deviation	Standard Error	Maximum	Minimum	Range
1	11	1.98	5.47	1.65	10.76	- 6.10	16.85
2	12	-0.74	10.75	3.10	14.33	-23.02	37.35
3	10	-0.31	7.85	2.48	8.54	-12.32	20.86
4	13	2.54	7.80	2.16	14.33	- 7.89	22.22
5	11	-2.26	9.26	2.79	10.88	-21.12	32.00
6	10	4.97	6.72	2.13	12.11	- 8.22	20.43
7	13	-2.65	10.66	2.96	11.88	-21.12	33.00
8	13	-1.14	7.43	2.06	10.33	-15.89	26.22
9	8	-1.40	8.69	3.07	12.79	-10.65	23.45
10	10	1.31	10.31	3.26	16.11	-13.24	29.35
11	11	-1.85	7.36	2.22	9.56	-12.32	21.88

In order to determine the effect of WISC Verbal IQ on the posttest, the following question was investigated: "Is there a significant difference in residual gain scores between E.M.H. students with 'low', 'middle', and 'high' WISC Verbal IQ scores?"

To answer this question, residual gain scores from the 97 students with WISC Verbal IQ data were blocked on three different ranges of IQ scores: 66 and less, 67 to 70, and 80 and above. An analysis of variance was computed on the blocked residual gain scores. Table 15 summarizes the results of the analysis of variance, indicating no significant differences between the three blocks.

TABLE 15
ANOVA, E.M.H. Experimental Group
Residual Gain Blocked on Verbal IQ

Source	d.f.	Mean Square	F Ratio
Between Groups	2	121.7180	1.8056
Within Groups	94	67.4115	

$F_{.05 (2,94)} = 2.71$, no significant difference

Table 16 summarizes the means and standard deviations for the residual gain scores blocked on WISC Verbal IQ.

TABLE 16
Mean Residual Gain Scores and Standard Deviations
E.M.H. Experimental Group
Blocked on WISC Verbal IQ

WISC Verbal IQ	N	Mean Residual Gain	Standard Deviation
80+	21	1.85	7.66
67-79	55	-0.02	8.49
≤66	21	-2.90	7.97

To determine the effect of WISC Performance IQ on posttest success, the following question was investigated: "Is there a significant difference in residual gain scores between students with 'low', 'middle', and 'high' WISC Performance IQ scores?"

To answer this question, the residual gain scores of the 97 E.M.H. students with WISC Performance IQ scores were blocked on three different ranges of IQ scores: 66 or less, 67 to 79, and 80 or higher. An analysis of variance was computed on the blocked residual gain scores. Table 17

summarizes the results of the analysis of variance, indicating no significant difference between the three blocks.

TABLE 17

ANOVA, E.M.H. Experimental Group
Residual Gain Blocked on Performance IQ

Source	d.f.	Mean Square	F Ratio
Between Groups	2	30.2643	0.4364
Within Groups	94	69.3573	

$F_{.05 (2,94)} = 2.71$, no significant difference

Table 18 summarizes the means and standard deviations for the residual gain scores blocked on WISC Performance IQ.

TABLE 18

Mean Residual Gain Scores and Standard Deviations
E.M.H. Experimental Group
Blocked on WISC Performance IQ

WISC IQ Performance	N	Mean Residual Gain	Standard Deviation
80+	44	0.25	7.72
67-79	34	0.01	9.59
66	19	-1.82	7.16

Since the multiple stepwise regression indicated that sex was a significant ($P < .05$) predictor of success on the 16-item instrument, and the table of correlation coefficients indicated that males were achieving a greater success than females, the following question was investigated: "Is there a significant difference in residual gain scores between male

and female E.M.H. students?"

To answer this question, residual gain scores were blocked on male and female cells and an analysis of variance was computed. Table 19 summarizes the results of the analysis of variance, indicating a significant ($P < .05$) difference between the two groups.

TABLE 19
ANOVA, E.M.H. Experimental Group

Source	d.f.	Mean Square	F Ratio
Between Groups	1	337.6349	
Within Groups	120	70.4016	4.7958*

*Significant at the .05 level, $F_{.05 (1,120)} = 3.92$

Table 20 summarizes the means and standard deviations for the residual gain scores for males and females.

TABLE 20
Male and Female Mean Residual Gain Scores
and Standard Deviations, E.M.H. Experimental Group

Sex	N	Mean	Standard Deviation
Males	78	1.25	8.29
Females	44	-2.22	8.57

Discussion

The results of the analyses of variance with residual gain scores blocked on 'low', 'middle', and 'high' ranges of WISC Verbal and Performance IQ Scores indicate that students with low verbal or performance scores are attaining the same level of success in the posttest instrument

as those students with high verbal or performance scores. From these results it can be inferred that the effect of verbal and/or performance IQ on success on the test instrument has been minimized.

The results of the analysis of variance on residual gain scores between males and females indicate that male E.M.H. students attain a higher level of success than female E.M.H. students. Since the residual gain score nullifies the regression effect, the results support the findings from the multiple stepwise regression that male E.M.H. students achieved a higher level of success on the posttest instrument than female E.M.H. students. In view of the younger ages, but higher measured I.Q. of the boys (see Table 3) the investigators suggest that an apparent differential criterion (or criteria) may be used to place boys in E.M.H. classes that is not applied to girls. This suggestion bears investigation.

In viewing the E.M.H. literature, nothing has been encountered, to date, concerning the ability of the E.M.H. student to perform the required tasks involved in choosing correct responses on a four-choice, multiple-choice type instrument. The fact that the E.M.H. student can handle this type of instrument, in the form used by the BSCS E.M.H. project, is extremely encouraging. It now seems as though this ability of the E.M.H. student has been underrated.

The student success attained on the preliminary test instrument justifies the expenditure of additional funds to develop additional multiple-choice items to help assess the achievement of E.M.H. students during the use of additional experimental materials in the coming year.

Objective Achievement Tests:

Analysis of Variance

To investigate the question, "Is there a significant difference between the experimental and control groups on the posttest?" an analysis of variance was computed on posttest scores. The pretest means for the experimental and control groups were 37.21 and 37.20, respectively, indicating that covariance analysis was not necessary. Table 21 summarizes the means and standard deviations of the posttest scores for the experimental and control groups.

TABLE 21
Posttest Means and Standard Deviations
for the
E.M.H. Experimental and Control Groups

Group	N	Mean	Standard Deviation
Experimental	122	44.557	9.744
Control	118	39.136	9.488

Table 22 summarizes the results of the analysis of variance.

TABLE 22
ANOVA, E.M.H. Experimental and Control Posttest Scores

Source	Sum of Squares	d.f.	Mean Square	F Ratio
Between Groups	1763.2545	1	1763.2545	19.0579**
Within Groups	22019.9289	238		
Total	23783.1834	239		

**Significant at the .005 level, $F_{.005 (1,238)} = 7.88$

Discussion

The results of the analysis of variance indicated that there was a significant difference ($P < .005$) between the experimental and control groups on posttest scores. On the basis of these results we conclude that the limited parts of the experimental package of materials assessed by the objective tests did have an effect on E.M.H. students. Eleven of the sixteen items measured baseline information, leaving only five items on which appreciable amounts of growth could be observed.

Objective Achievement Tests:

Items and Objectives

The section that follows presents a detailed account of the test items in the context of the objectives to which they were judged to be related. The items were intended to sample, directly and indirectly, the extent to which students could perform the objective prior to instruction, the extent to which instruction could effect student responses to the items, and by inference, the extent to which students would perform the objectives after instruction.

The curriculum development staff assumed that the E.M.H. curriculum does not currently include any study of the internal human body. Therefore, no data were gathered on the curriculum of control group teachers in the period between pretest and posttest administration.

Items were prepared to assess each objective, randomly assigned to Form A and B, and then mixed and randomly assigned a numerical position with each test form.

The analysis below retrieved each item pair, one from Form A and one

from form B. The item number indicates the position of each item within the thirty-item test. Twenty-eight items were rejected from further analysis after a review of item analysis data and a critique of each item by the Evaluation Committee.

Item responses and biserial correlations for pretest and posttest for both experimental and control groups are provided in Table 23. Tables 24 to 40 were constructed by computing the difference in the proportion (percentage) of students selecting each response between the pretest and the posttest. Weighted averages were calculated for the combined responses for each matching pair.

TABLE 23

Item Responses and Biserial Correlations for
Experimental and Control Groups, Forms A and B

Student Group	Item Numbers	Pretest								Posttest							
		N	Percent of N							N	Percent of N						
			A	B	C	D	M*	O*	r_b		A	B	C	D	M	O	r_b
Experimental	A-1	68	7	2	62	28	0	2	.38	54	4	0	74	21	2	0	.64
	B-1	54	48	13	33	4	0	2	.31	68	62	6	31	0	0	2	.32
Control	A-1	56	10	5	46	38	2	0	.18	62	6	2	66	26	0	0	.54
	B-1	62	60	10	24	5	0	2	.49	56	39	7	52	0	0	2	.36
Experimental	A-6	68	22	0	44	13	21	0	.29	54	44	2	33	6	15	0	.60
	B-7	54	17	17	43	4	20	0	.53	68	6	15	50	28	0	2	.55
Control	A-6	56	23	4	57	14	2	0	.46	62	39	0	50	8	3	0	.55
	B-7	62	7	19	61	11	2	0	.46	56	4	23	55	16	0	2	.60
Experimental	A-8	68	91	2	7	0	0	0	.07	54	98	0	2	0	0	0	1.04
	B-8	54	0	0	0	96	4	0	.28	68	2	0	0	96	3	0	.20
Control	A-8	56	82	2	16	0	0	0	.46	62	97	0	3	0	0	0	.59
	B-8	62	100	0	0	1	0	0	.00	56	4	4	0	93	0	0	1.00
Experimental	A-4	68	2	2	96	0	2	0	.40	54	0	2	94	0	4	0	.54
	B-14	54	2	0	0	96	2	0	.70	68	3	0	3	94	0	0	.63
Control	A-4	56	2	5	82	2	9	0	.68	62	0	0	98	2	0	0	-.09
	B-14	62	2	2	3	92	2	0	.52	56	2	5	5	86	0	2	.78
Experimental	A-13	68	3	0	75	22	0	0	.44	54	2	0	74	22	2	0	.73
	B-4	54	78	0	2	15	6	0	.59	68	82	0	2	16	0	0	.42
Control	A-13	56	4	0	77	16	2	2	.42	62	2	0	84	15	0	0	.39
	B-4	62	81	5	2	11	0	2	.35	56	80	2	0	16	0	2	.26
Experimental	A-3	68	9	18	65	6	2	2	.60	54	0	17	72	4	6	2	-.06
	B-15	54	9	22	6	50	11	2	.37	68	3	31	4	62	0	0	.51
Control	A-3	56	5	25	57	9	0	4	.32	62	2	7	79	13	0	0	.70
	B-15	62	5	26	8	61	0	0	.14	56	9	27	4	59	0	2	.42
Experimental	A-10	68	3	16	38	41	2	0	.34	54	7	61	21	9	2	0	.57
Control	A-10	56	16	11	32	39	0	2	.12	62	5	40	27	27	0	0	.60
Experimental	A-20	68	15	6	66	7	0	6	.41	54	6	0	91	4	0	0	.61
	B-10	54	6	7	80	4	4	0	.29	68	3	2	93	3	0	0	.88
Control	A-20	56	9	9	75	7	0	0	.61	62	11	2	77	10	0	0	.56
	B-10	62	10	2	79	7	3	0	.50	56	7	5	79	9	0	0	.68
Experimental	A-12	68	9	3	10	77	2	0	.47	54	4	2	4	89	2	0	.58
	B-27	54	7	4	78	4	6	2	.45	68	4	6	87	3	0	0	.82
Control	A-12	56	5	5	7	77	2	4	.59	62	7	7	2	86	0	0	.70
	B-27	62	15	7	74	3	0	2	.50	56	14	4	75	5	0	2	.59
Experimental	A-5	68	38	50	0	12	0	0	.56	54	50	30	7	11	2	0	.43
	B-9	54	17	52	13	13	6	0	.44	68	65	13	7	15	0	0	.69
Control	A-5	56	18	57	7	13	0	5	.69	62	52	36	2	7	3	2	.53
	B-9	62	37	55	3	5	0	0	.57	56	32	43	14	11	0	0	.61
Experimental	A-9	68	0	13	59	15	13	0	.50	54	2	11	80	6	2	0	.72
	B-17	54	22	0	56	17	6	0	.30	68	15	0	69	16	0	0	.55
Control	A-9	56	4	25	48	23	0	0	.49	62	0	10	79	11	0	0	.39
	B-17	62	34	0	50	15	0	2	.25	56	32	2	59	5	0	2	.30

TABLE 23 (continued)

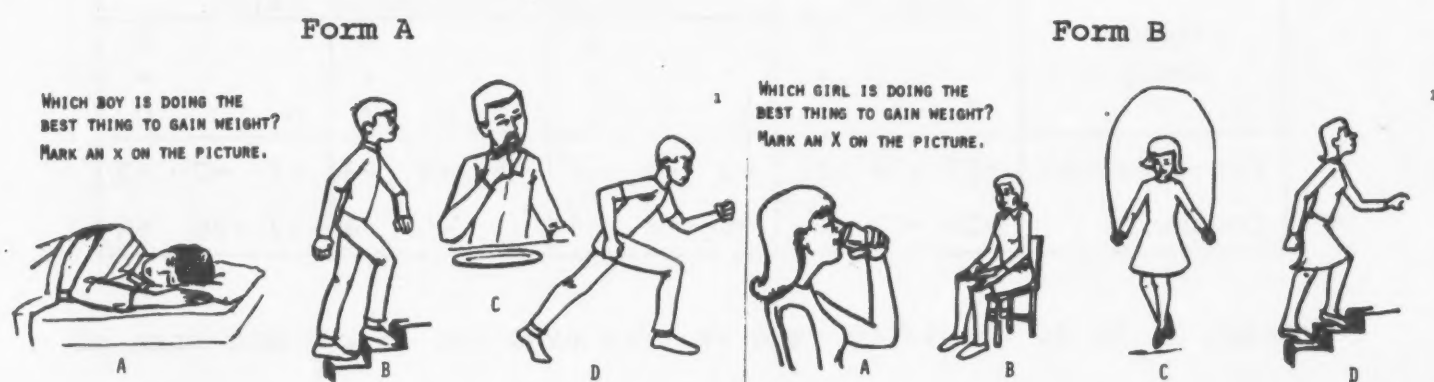
Item Responses and Biserial Correlations for
Experimental and Control Groups, Forms A and B

Student Group	Item Numbers	Pretest								Posttest							
		N	Percent of N							N	Percent of N						
			A	B	C	D	M*	O*	r_b		A	B	C	D	M	O	r_b
Experimental	A-7	68	16	10	9	63	0	2	.48	54	7	0	6	85	2	0	.40
	B-6	54	7	72	13	6	0	2	.64	68	0	91	2	3	3	2	.54
Control	A-7	56	9	9	13	66	0	4	.48	62	7	0	2	92	0	0	.42
	B-6	62	7	82	7	5	0	0	.37	56	14	59	13	14	0	0	.46
Experimental	A-18	68	13	7	77	3	0	0	.44	54	17	7	70	0	6	0	.73
	B-24	54	2	4	78	17	0	0	.39	68	0	3	81	16	0	0	.70
Control	A-18	56	5	2	86	4	2	2	.49	62	37	3	58	2	0	0	.14
	B-24	62	3	2	73	21	0	2	.42	56	7	5	70	13	2	4	.12
Experimental	A-19	68	16	66	6	9	2	2	.56	54	13	72	6	7	2	0	.70
	B-22	54	4	11	50	17	19	0	.55	68	10	4	79	6	0	0	.90
Control	A-19	56	27	59	4	9	0	2	.45	62	26	57	5	13	0	0	.53
	B-22	62	8	3	68	21	0	0	.22	56	7	7	46	38	0	2	.34
Experimental	B-13	54	9	6	48	33	4	0	-.04	68	7	22	60	10	0	0	.27
Control	B-13	62	15	11	60	13	0	2	.32	56	16	5	39	34	2	4	-.29
Experimental	A-15	68	31	10	18	32	0	9	.43	54	11	33	17	39	0	0	.58
	B-25	54	37	15	26	20	2	0	.49	68	24	18	15	43	2	0	.36
Control	A-15	56	20	5	21	52	0	2	-.07	62	8	8	10	74	0	0	.73
	B-25	62	37	10	18	31	0	5	.35	56	29	27	25	18	0	2	.32
Experimental	A-26	68	35	12	27	24	2	2	.23	54	30	9	20	35	6	0	.43
	B-26	54	15	7	43	28	7	0	.26	68	13	19	24	41	0	3	.40
Control	A-26	56	29	18	18	34	0	2	.37	62	37	18	16	29	0	0	.50
	B-26	62	15	11	45	26	0	3	.68	56	25	5	43	27	0	0	.34

*M represents multiple responses that included the correct responses and O represents omissions.

Objective 1. Students will associate foods with generalized body needs.

Three pairs of items were designed to sample achievement of this objective. One pair (A-2, B-2) was discarded as the items were judged to be irrelevant to the objective.



Item pair A-1, B-1 was judged to provide baseline data, that is, data that when grouped with other baseline items would provide a profile of entry-level behaviors for a class group. Baseline data items are considered to be useful for establishing a point of reference for individuals and groups but to be of less value in providing information on the achievement of students as a result of the instructional program. This item pair was judged to be lacking in parallel construction as one item provides eating as the best choice and the other provides drinking. Item B-1 should be revised to show eating.

Table 24 shows that the raw score gains made by the experimental group resulted in shifts from selecting exercise and rest to selecting eating (drinking) as the best way to gain weight. A shift from rest and light exercise to heavy exercise occurred in the control group.

TABLE 24

Pretest to Posttest Response Changes*
for Item Pair A-1, B-1
(The response choice for A-1 is cited first.)

Student Group	Percent Change, Pretest to Posttest											
	Correct Choice			Parallel Distractor Pairs								
	C			A			B			D		
	+			+			+			+		
	C	A	A	A	B	B	B	D	D	D	C	C
Experimental	+12	+14	+11	-3	-7	-5	-2	-4	-3	-7	-2	-3
Control	+20	-21	0	-4	-3	-3	-3	-5	-4	-12	+28	+8

*Tables 24 to 40 should be read as, for example, 12 percent more of the experimental group selected response C in item A-1 on the posttest as compared to the pretest.

Form A

IF YOU WERE GOING TO GO ON A SPACE TRIP THAT WOULD LAST FOR DAYS AND DAYS, WHICH OF THESE WOULD YOU NEED MOST?

MARK AN X ON THIS PICTURE.



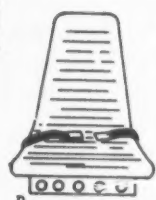
A



B



C



D

Form B

IF YOU WERE GOING TO GO ON A SPACE TRIP THAT WOULD LAST FOR DAYS AND DAYS, WHICH OF THESE DO YOU NEED MOST?

MARK AN X ON THE PICTURE.



A



B



C



D

Gains (see Table 25) were made by both the experimental and control groups on item pair A-6, B-7. The gain for the control group can be accounted for by a loss in selecting music (A-6-B, B-7-A). Net changes occurred in the major distractor pair (A-6-C, B-7-B) for the experimental group, but not the control group. Choices B and D of item B-7 may not have been perceived as being related to space flight as much as choices C and D of item A-6. The shift to food (choices A and C) by the experimental group may be attributed to instruction. Teacher comments indicated student difficulty in identifying B-7-D.

The evaluation committee recommended that the art work of B-7 be clarified and that the helmet in A-6-C be replaced by gloves, to better equalize the distractor pair. The items are considered to provide baseline, rather than achievement data.

TABLE 25

Pretest to Posttest Response Changes
for Item Pair A-6, B-7
(The response choice for A-6 is cited first.)

Student Group	Percent Change, Pretest to Posttest											
	Correct Choice			Parallel Distractor Pairs								
	A			B			C			D		
	+			+			+			+		
	A	C	C	B	A	A	C	B	B	D	D	D
Experimental	+22	+7	+16	+2	-11	+4	-11	-2	-9	-7	+24	+9
Control	+16	-6	+5	-4	-3	-4	-7	+4	0	-6	+5	0

Objective 2. Students will associate food types with plant and animal sources.

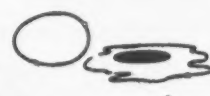
Form A

WHICH FOOD COMES FROM ANIMALS?
MARK AN X ON THE FOOD.



Form B

WHICH FOOD COMES FROM PLANTS?
MARK AN X ON THE PLANT FOOD.



One pair of items, A-8 and B-8 was designed to sample achievement of this objective. Item A-8 required the selection of one animal-source food from three foods of plant origin and item B-8 required selection of one plant-source food from three animal-source foods. Success on these items indicates that students have prior knowledge of this objective, although a single pair of items does not indicate the range of their knowledge.

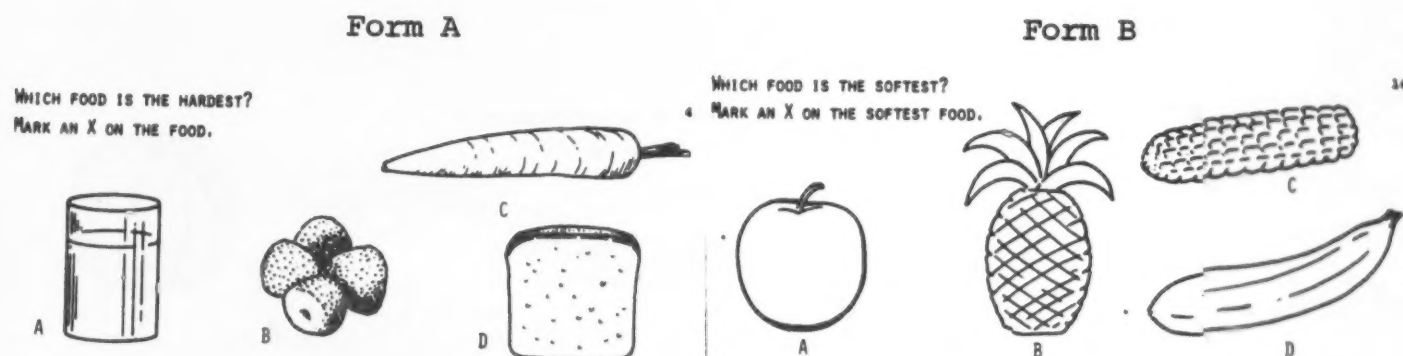
TABLE 26

Pretest to Posttest Response Changes
for Item Pair A-8, B-8
(The response choice for A-8 is cited first.)

Student Group	Percent change, Pretest to Posttest											
	Correct Choice			Other Choices*								
	A	D	D	C	A	C	B	B	B	D	C	C
Experimental	+7	0	+2	-5	+2	-3	-2	0	-1	0	0	0
Control	+15	-7	+4	-13	+4	-4	-2	+4	+1	0	0	0

*There are no corresponding responses on this item. M_A and M_B represent multiple responses that included the correct response.

Objective 3. Students will distinguish differences in physical characteristics of foods.



One item pair, A-4 and B-14 was designed to secure baseline data on this objective. High initial achievement (Experimental group, Pretest A = 96%, B = 94%; Control group, Pretest A = 82%, B = 98%) left no growth potential for this item pair. Students have prior knowledge that enabled them to select the hardest foods from two arrays of four foods each. However, no assessment was made of the consequences of hardness in chewing and eating.

TABLE 27

Pretest to Posttest Response Changes
for Item Pair A-4, B-14
(The response choice for A-4 is cited first.)

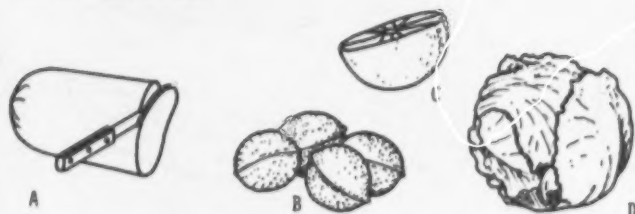
Student Group	Percent Change, Pretest to Posttest											
	Correct Choice			Parallel Distractor Pairs								
	C	D	D	D	C	C	B	A	A	A	B	B
Experimental	-2	-2	-2	0	+3	+2	0	+1	0	-2	0	-1
Control	+16	-6	+5	0	+2	+2	-5	0	-3	-2	+3	0

*M_A and M_B represent multiple responses that included the correct response.

Objective 4. Students will identify water as a substance present in many foods.

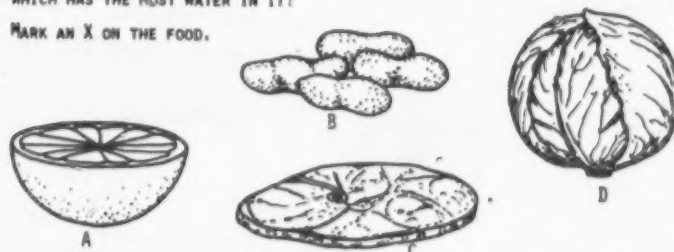
Form A

WHICH FOOD HAS THE MOST WATER IN IT?
MARK AN X ON THE FOOD.



Form B

13. WHICH HAS THE MOST WATER IN IT?
MARK AN X ON THE FOOD.



One pair of items, A-13 and B-4 was designed to secure baseline data on this objective. In both pretesting and posttesting, students chose citrus fruit to lettuce or cabbage by a ratio of about 3:1 or 4:1. Citrus fruit was the test writers' choice. This type of item raised the issue of the degree of discrimination desirable for this student population. Also, the item applies to the objective in a tangential, rather than direct way. The item does assess what might be called "common sense". Seventy-five to 80 percent of the student population agreed with the test writers. A factual check indicates that in both instances, "common sense" is inaccurate. Lettuce and cabbage contain from 92 to 95 percent water, whereas oranges, for example, contain only 85 percent water. An examination of comparative changes between experimental and control groups supports the interpretation that instruction had no affect on response selection for this item.

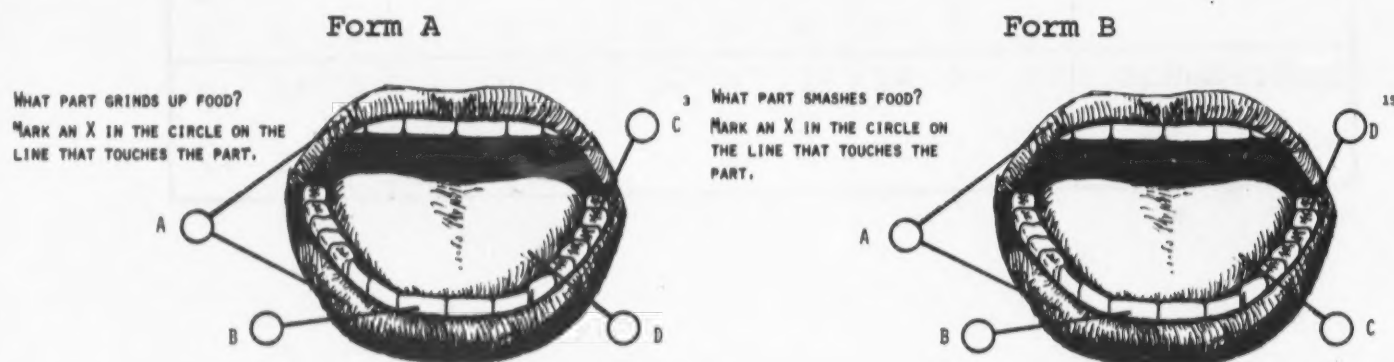
TABLE 28

Pretest to Posttest Response Changes
for Item Pair A-13, B-4
(The response choice for A-13 is cited first.)

Student Group	Percent Change, Pretest to Posttest											
	Correct Choice			Parallel Distractor Pairs								
	C +			A +			B +			D +		
	C	A	A	A	C	C	B	B	B	D	D	D
Experimental	-1	+4	+2	-1	0	-1	0	0	0	0	+1	0
Control	+7	-1	+3	-2	-2	-2	0	-3	-2	-1	+5	+2

Objective 5. Students will observe and describe the structure and functions of mouth parts.

Four item pairs were developed to assess this objective. Item pair A-16, B-3 was judged to be nonparallel and to imply an exclusive bite function to a particular group of teeth. These items were omitted from further analysis.



Item pair A-3, B-15 was designed to test recall knowledge of the function of the molars. Both experimental and control groups gained on posttest scores on this item. Seventy percent of the control group responded correctly on the average of the posttest pair, compared to 66 percent of the experimental group. The implication of singular functions may contribute to the weakness of this item pair.

Instruction cannot be interpreted to have contributed pretest to posttest gains in view of the gains made by the control group.

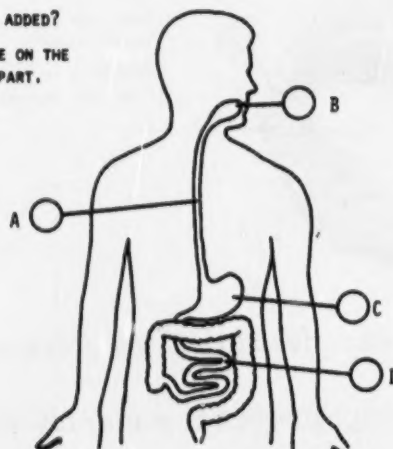
TABLE 29

Pretest to Posttest Response Changes
for Item Pair A-3, B-15
(The response choice for A-3 is cited first.)

Student Group	Percent Change, Pretest to Posttest											
	Correct Choice			Parallel Distractor Pairs								
	C			A			B			D		
	+			+			+			+		
	C	D	D	A	A	A	B	B	B	D	C	C
Experimental	+7	+12	+8	-9	-6	-7	-1	+9	+5	-2	-2	-2
Control	+22	-2	+11	-3	+4	0	-18	+1	-10	+4	-4	+1

Form A

IN WHAT PART IS SALIVA ADDED?
MARK AN X IN THE CIRCLE ON THE
LINE THAT TOUCHES THE PART.



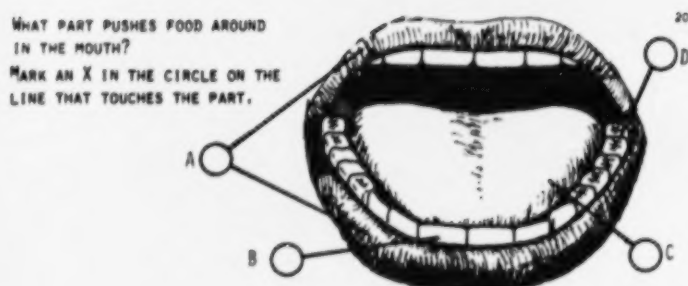
Item pair A-10, B-19 was judged to be nonparallel. Item B-19 assessed a concept not included in the instructional program and was omitted from further analysis. Item A-10 is a recall item that assessed the students' knowledge of where saliva is added. The raw score gains made by the experimental group indicated a large shift from selecting the stomach and intestine to the mouth, after instruction.

TABLE 30

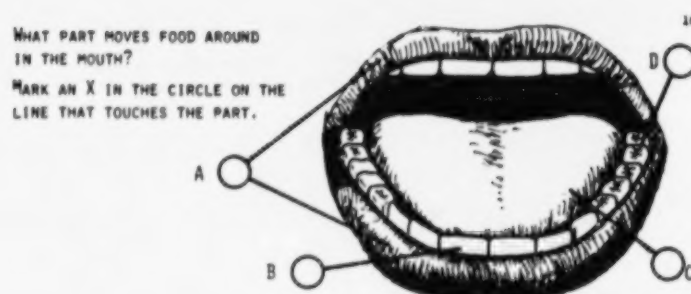
Pretest to Posttest Response Changes
for Item A-10

Student Group	Percent Change, Pretest to Posttest			
	Correct Choice	Other Choices		
		B	A	C D
Experimental	+45		+4	-17 -32
Control	+29		-11	-5 -12

Form A



Form B



Item pair A-20, B-10 was designed to obtain information on tongue function. Experimental group pretest success was high (A-20, 66 percent and B-10, 80 percent), but gain from pretest to posttest with 20 percent more students selecting the correct response in contrast to only 1 percent more for the control group, may be attributed to achievement based on instruction.

Teachers commented that the striking appearance of the mouth tended to distract students. They recommended a change in this artwork for all the items in which it was used.

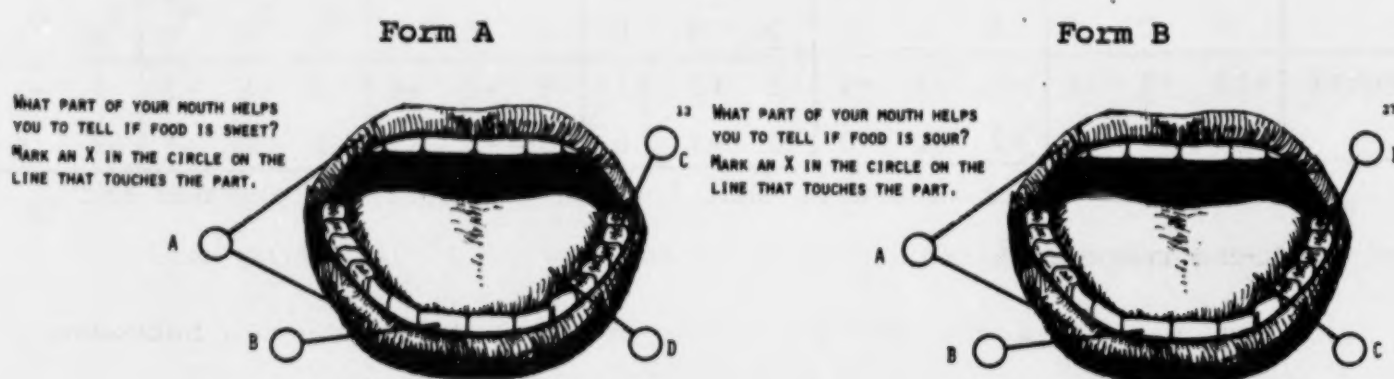
TABLE 31

Pretest to Posttest Response Changes
for Item Pair A-20, B-10
(The response choice for A-20 is cited first.)

Student Group	Percent Change, Pretest to Posttest											
	Correct Choice			Parallel Distractor Pairs								
	C			A			B			D		
	C	C	C	A	A	A	B	B	B	D	D	D
Experimental	+25	+13	+20	-9	-3	-7	-6	-5	-5	-3	-1	-3
Control	+2	0	+1	+2	-3	-1	-7	+3	-2	+3	+2	+3

Objective 6. Students will relate functions of teeth, tongue, and saliva in the mouth to characteristics of foods.

Two item pairs were designed to assess student performance for this objective. Item pair A-23, B-20 was judged to be faulty in construction in implying that only one mouth part is used to eat pudding. This item pair was omitted from further analysis.



Item pair A-12, B-27 was designed to assess tongue function in relation to food characteristics. The experimental group shifted (pretest to posttest) from two distractors and multiple responses to the correct response in the 11 percent gain in the percentage of students selecting the correct response. Approximately half of the gain from pretest to posttest for the control group was attained by a shift from multiple responses and omissions. This item pair is considered to provide baseline data since over 75 percent of both experimental and control groups responded correctly to the item on the pretest.

TABLE 32

Pretest to Posttest Response Changes
for Item Pair A-12, B-27
(The response choice for A-12 is cited first.)

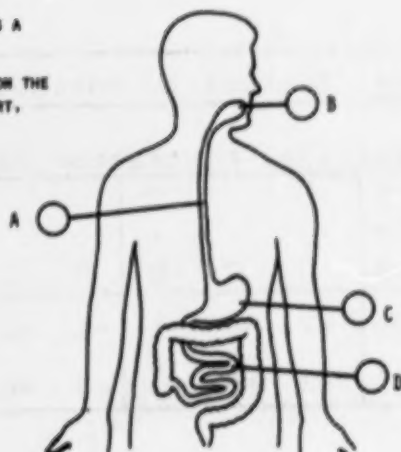
Student Group	Percent Change, Pretest to Posttest																	
	Correct Choice			Parallel Distractor Pairs														
	D +			A +			B +			C +			M_A^* +			O_A^+ +		
	D	C	C	A	A	A	B	B	B	C	D	D	M_A^*	M_B	M_B^+	O_A	O_B	O_B^+
Experimental	+12	+9	+11	-5	-3	-4	-1	+2	+1	-6	-1	-4	0	-6	-3	0	-2	-1
Control	+9	-1	+6	+2	-1	0	+2	-3	0	-5	+2	-2	-2	0	-1	-4	0	-2

* M_A and M_B represent multiple responses that include the correct response and O_A and O_B represent omitted responses.

Objective 7. Students will describe the location and function of the esophagus.

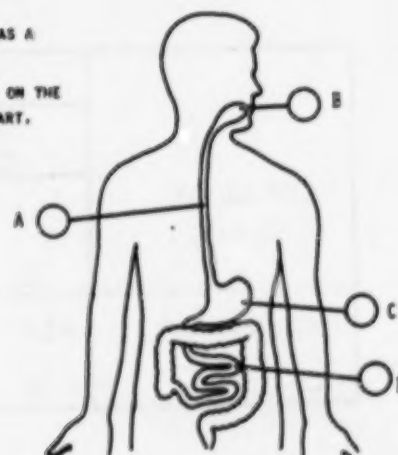
Form A

WHAT PART IS USED JUST AS A
PASSAGE WAY FOR FOOD?
MARK AN X IN THE CIRCLE ON THE
LINE THAT TOUCHES THE PART.



Form B

WHAT PART IS USED JUST AS A
PASSAGE WAY FOR FOOD?
MARK AN X IN THE CIRCLE ON THE
LINE THAT TOUCHES THE PART.



Only one item pair, A-5, B-9, was designed to assess this objective. These items were identical in construction. Use of the word "just" in the stem was questioned by the evaluation committee. A key word such as this should be avoided in future item development if possible.

This item is clearly an achievement item. About 25 percent of both experimental and control groups responded successfully on the pretest.

A gain of 31 percent on the posttest by the experimental group, in contrast to a gain of 15 percent in the control group is the basis for judging this item to assess achievement based directly on instruction. The stem is, however, a factual recall item, and of minor importance to the total instructional program.

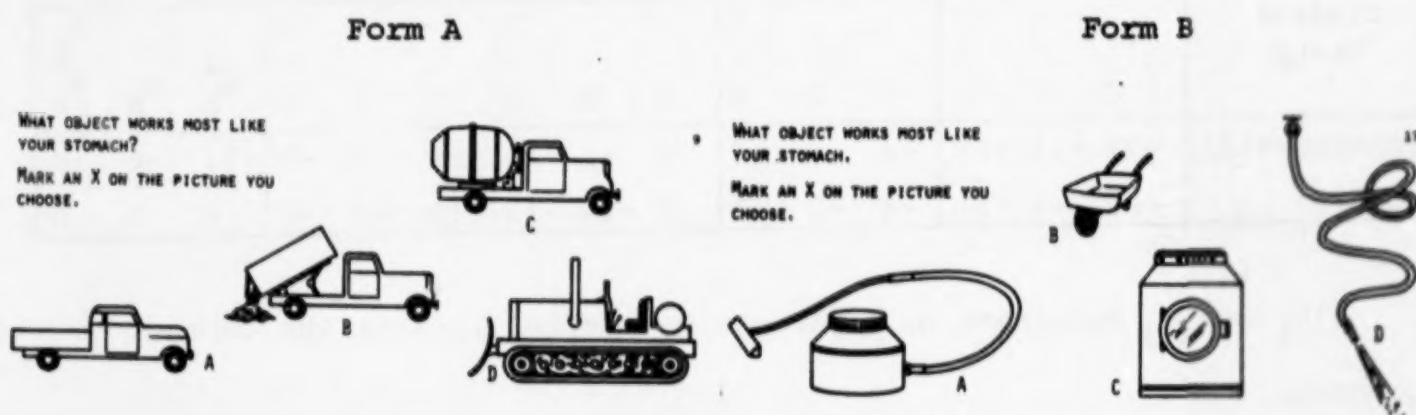
TABLE 33

Pretest to Posttest Response Changes
for Item Pair A-5, B-9
(The response choice for A-5 is cited first.)

Student Group	Percent Change, Pretest to Posttest											
	Correct Choice			Parallel Distractor Pairs								
	A			B			C			D		
	+			+			+			+		
	A	A	A	B	B	B	C	C	C	D	D	D
Experimental	+12	+48	+31	-20	-39	-30	+7	-6	+1	-1	+2	+1
Control	+34	-5	+15	-21	-12	-17	-3	+9	+3	-4	+6	0

Objective 8. Students will relate stomach functions to changes in food form.

Two item pairs were designed to provide data on student performance on this objective. Item pair A-17, B-5 was not related to the objective and was omitted from further analysis.



Item pair A-9, B-17 was judged to assess the relational level of cognitive difficulty and to provide evidence of conceptual thinking.⁷ Both experimental and control groups showed gains on the posttest in the proportion of students selecting the correct response.

A flat statement that the experimental materials contributed to gains is contradicted by the strong gains by the control group. The strongest gains have been made consistently by control classes taking Form A as a posttest (see Tables 24 to 40), which will be accounted for in the summary of this report.

⁷See pages 3, 4.

TABLE 34

Pretest to Posttest Response Changes
for Item Pair A-9, B-17
(The response choice for A-9 is cited first.)

Student Group	Percent Change, Pretest to Posttest														
	Correct Choice			Parallel Distractor Pairs											
	C			A			B			D			M_A^+		
	C	C	C	A	B	B	B	D	D	D	A	A	M_A^*	M_B	M_B^+
Experimental	+21	+13	+16	+2	0	+1	-2	-1	-1	-9	-7	-7	-11	-6	-8
Control	+31	+9	+21	-4	+2	-1	-15	-10	-12	-12	-2	-8	0	0	0

* M_A and M_B represent multiple responses that include the correct response.

Objective 9. Students will locate and describe external evidence of circulation.

Form A

Form B

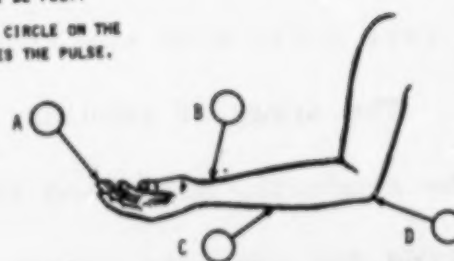
WHERE CAN A PULSE BE FELT?

MARK AN X IN THE CIRCLE ON THE ARROW THAT TOUCHES THE PULSE.



WHERE CAN A PULSE BE FELT?

MARK AN X IN THE CIRCLE ON THE ARROW THAT TOUCHES THE PULSE.



Only one item pair, A-7, B-6, was developed to assess this objective. This item pair was designed to provide background data and possible achievement data on where a pulse can be detected on the body. Approximately two-thirds of both experimental and control group students responded correctly to these items on the pretest.

Posttest gains by the experimental group is attributed to instruction as there were losses on all distractors on both item forms that account for this increased proportion of students responding correctly.

TABLE 35

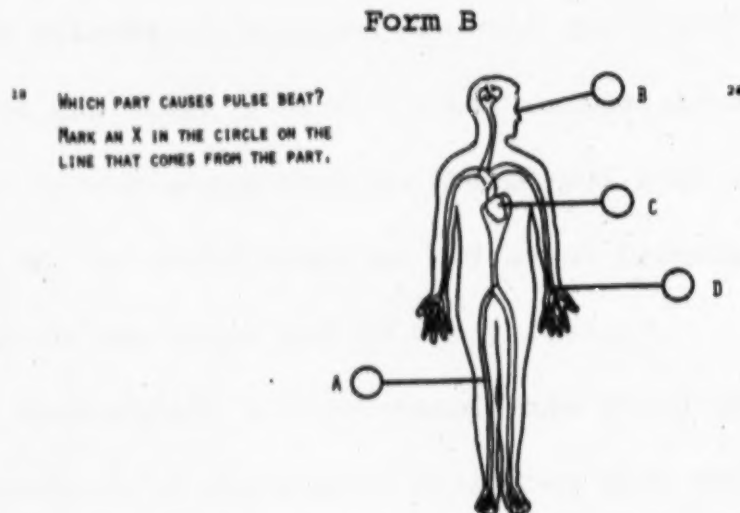
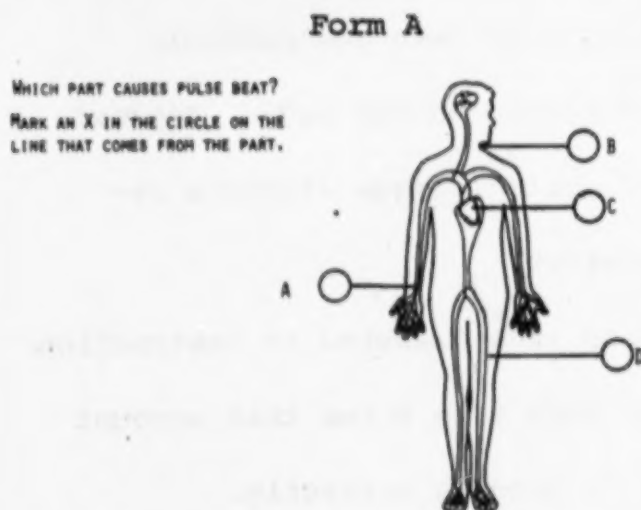
Pretest to Posttest Response Changes
for Item Pair A-7, B-6
(The response choice for A-7 is cited first.)

Student Group	Percent Change, Pretest to Posttest											
	Correct Choice			Other Choices*								
	D +			B +			A +			C +		
	D	B	B	B	D	D	A	C	C	C	A	A
Experimental	+22	+19	+21	-10	-3	-6	-9	-11	-11	-3	-7	-5
Control	+26	-23	+2	-9	+9	0	-2	+6	+2	-11	+7	-2

*There are no corresponding responses on these items.

Objective 11. Students will associate heart action with heart sounds and pulse.

Five item pairs were designed to assess student performance on this objective. Three of the five pairs were omitted from further analysis. Item pair A-11, B-16 was rejected because of the negative working of the stem. The array of tabular information was new and apparently confusing to the students, according to teacher reports. Item pair A-24, B-18 was rejected for the same reason. Item pair A-28, B-11 was judged to yield no useful information.



Item pair A-18, B-24 was designed to determine baseline relationships between heart and pulse beat. Failure of the experimental group to gain in this item pair can be attributed to a failure of instruction. The causal relation, as measured by this item, was clearly not learned. The shift from pretest to posttest was to select where pulse is taken.

TABLE 36

Pretest to Posttest Response Changes
for Item Pair A-18, B-24
(The response choice for A-18 is cited first.)

Student Group	Percent Change, Pretest to Posttest											
	Correct Choice			Parallel Distractor Pairs								
	C			A			B			D		
	C	C	C	A	D	D	B	B	B	D	A	A
Experimental	-7	+3	-1	+4	-1	+1	0	-1	-1	-3	-2	-3
Control	-28	-3	-15	+32	-8	+13	+1	+3	+2	-2	+4	+1

Form A

WHICH OF THESE DOES THE HEART
MOVE: AIR, BLOOD, SALIVA,
STOMACH JUICE.

MARK AN X ON YOUR CHOICE.

AIR	BLOOD	SALIVA	STOMACH JUICES
A	B	C	D

Form B

19 WHICH OF THESE DOES THE HEART
MOVE: STOMACH JUICES, SALIVA,
BLOOD, AIR.

MARK AN X ON YOUR CHOICE.

STOMACH JUICES	SALIVA	BLOOD	AIR
A	B	C	D

Item pair A-19, B-22 was designed to measure student association of heart and blood movement. Somewhat less than two-thirds of both experimental and control groups responded correctly on the pretest to this item pair. The experimental group made posttest gains in correct response; the control group lost. For both groups, the heart moving air was the most popular distractor. Substituting the word "pumps" for "moves" may clarify this choice, since this item pair does not enable investigators to distinguish the knowledge of air (or oxygen) distribution from blood distribution by heart action.

TABLE 37

Pretest to Posttest Response Changes
for Item Pair A-19, B-22
(The response choice for A-19 is cited first.)

Student Group	Percent Change, Pretest to Posttest											
	Correct Choice			Parallel Distractor Pairs								
	B			A			C			D		
	B	C	C	A	D	D	C	B	B	D	A	A
Experimental	+6	+29	+17	-3	-11	-7	0	-7	-3	-2	+6	+2
Control	-2	-22	-12	-1	+17	+8	+1	+4	+3	+4	-1	+2

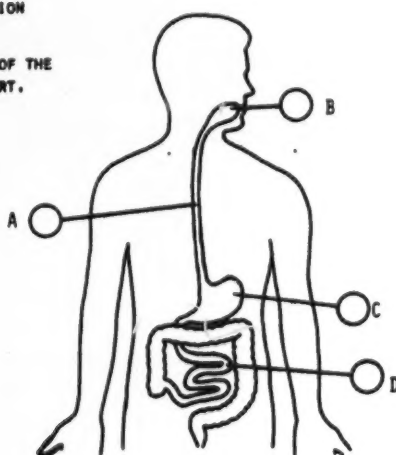
Objective 14. Students will infer a relationship between food and blood and describe barriers between them.

Three item pairs were designed to sample student performance for this objective. Item pair A-29, B-29 was omitted from further analysis in that the items were incomprehensible to the students. Item pair A-27, B-12 was also omitted from further analysis. Examination of stem data by the evaluation committee led them to judge symbol-object confusion that made the items incomprehensible.

Item A-25 of item pair A-25, B-13 was omitted because the idea of where digestive process stops was not taught and is irrelevant to the objective.

Form B

IN WHAT PART DOES DIGESTION
START?
MARK AN X IN THE CIRCLE OF THE
LINE THAT TOUCHES THE PART.



Item B-13 was difficult for both experimental and control groups and is therefore considered to be a new idea for these students. Gain on the posttest by the experimental group and loss by the control group is judged to attribute the gain to instruction. Selection of response D by both experimental and control groups on both pretest and posttest does indicate general knowledge of the location of digestion in the body.

TABLE 38

Pretest to Posttest Response Changes
for Item B-13

Student Group	Percent Change, Pretest to Posttest			
	Correct Choice	Other Choices*		
	B	A	C	D
Experimental	+16	-2	+12	-23
Control	-6	+7	+21	+21

*There are no corresponding responses
on this item.

Objective 16. Students will associate circulation with distribution of food.

Four item pairs were constructed to assess student performance on this objective. Item pair A-21, B-30 was omitted from further analysis as the items were judged to be inconsistent with the materials. The materials used the conception of liquifying food rather than reducing large particles to small particles in the digestive process. Item pair A-30, B-28 was also omitted for analysis on the same basis for the item pair above.

Form A

Form B

SUGAR IN THE BLOOD IN YOUR TOE
WAS ONCE A PART OF: HEART, BREAD,
STOMACH, BLOOD VESSEL.
MARK AN X ON YOUR CHOICE.

HEART
ASLICE OF BREAD
BSTOMACH
CBLOOD VESSEL
D

15 SUGAR IN YOUR ARM WAS ONCE A
PART OF: BLOOD VESSEL, STOMACH,
HEART, CRACKER.
MARK AN X ON YOUR CHOICE.

BLOOD VESSEL
ASTOMACH
BHEART
CCRACKER
D

Item pair A-15, B-25 was judged to assess student performance at the level of application. Although the experimental group made a large post-test gain, less than fifty percent of the students were able to select the correct choice on the posttest.

TABLE 39

Pretest to Posttest Response Changes
for Item Pair A-15, B-25
(The response choice for A-15 is cited first.)

Student Group	Percent Change, Pretest to Posttest											
	Correct Choice			Parallel Distractor Pairs								
	B			A			C			D		
	B	C	C	A	C	C	C	D	D	D	A	A
Experimental	+23	+23	+25	-20	-11	-16	-1	+3	+1	+7	-13	-3
Control	+3	-15	-6	-12	0	-6	-11	+17	+3	+22	-8	+9

Form A

WHICH MATERIAL CAN BE FOUND IN
BLOOD: SUGAR, STARCH, SALIVA,
STOMACH JUICE.
MARK AN X ON YOUR CHOICE.

SUGAR	STARCH	SALIVA	STOMACH JUICE
A	B	C	D

Form B

26 WHICH MATERIAL CAN BE FOUND IN
BLOOD: SALIVA, STARCH, SUGAR,
STOMACH JUICE?
MARK AN X ON YOUR CHOICE.

SALIVA	STARCH	SUGAR	STOMACH JUICE
A	B	C	D

Item pair A-26, B-26 was judged to assess student performance at the level of application. Both experimental and control groups responded to these items at about the 40 percent level on the pretest. The experimental group showed a marked shift to choices D,D on the posttest. The materials emphasized the liquification of food by stomach juices and the movement of liquid into the blood. The students learned from instruction, but is this a desired learning or should the materials be revised to clarify substances that do and do not move into the blood?

TABLE 40

Pretest to Posttest Response Changes
for Item Pair A-26, B-26
(The response choice for A-26 is cited first.)

Student Group	Percent Change, Pretest to Posttest											
	Correct Choice			Parallel Distractor Pairs								
	A			B			C			D		
	A	C	C	B	B	B	C	A	A	D	D	D
Experimental	-5	-19	-12	-3	+12	+5	-7	-2	-6	+11	+13	+12
Control	+8	-2	+3	0	-6	-2	-7	-2	-4	-5	+1	-2

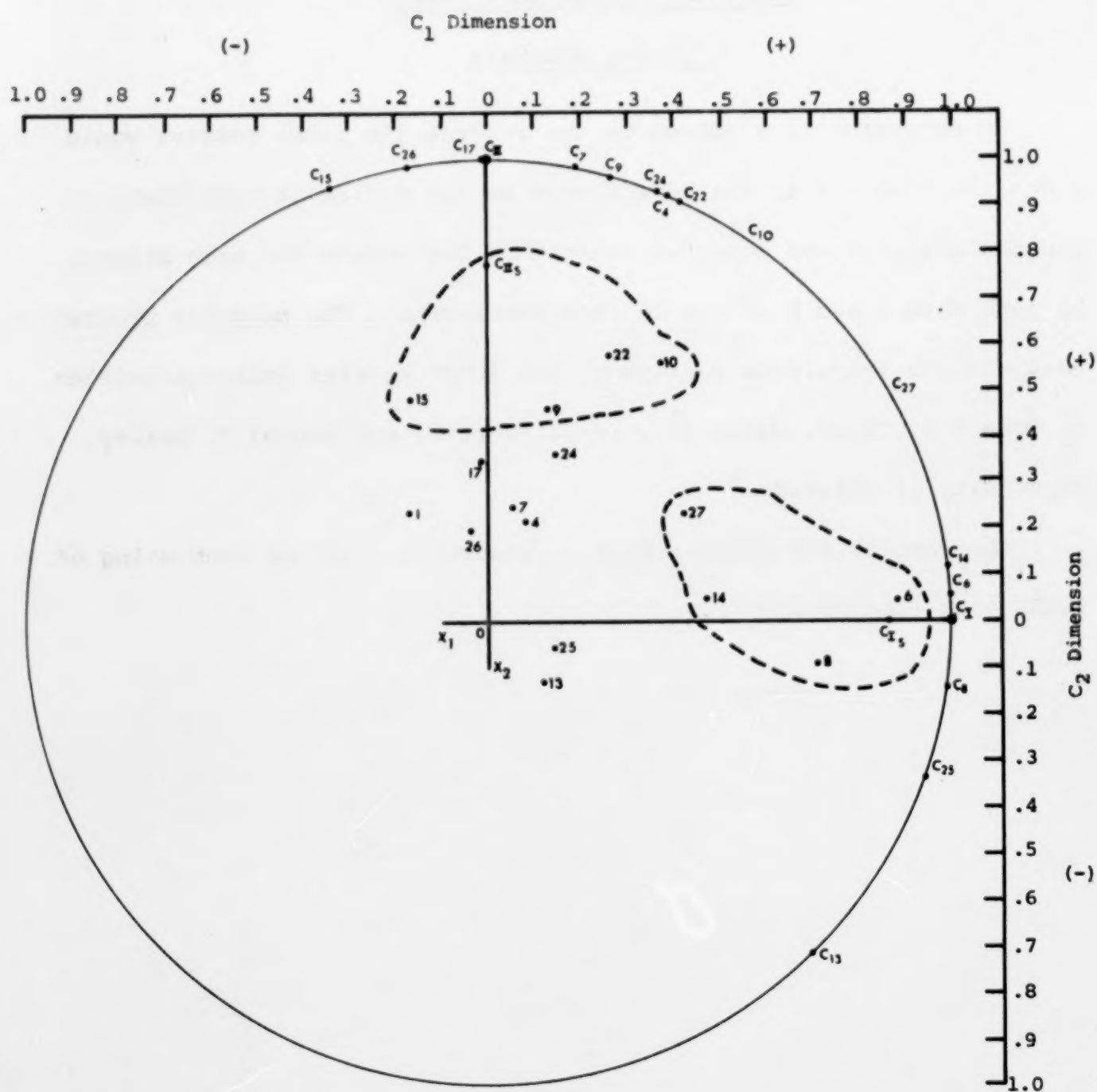
Objective Achievement Tests:

Cluster Analysis

To determine if a subset of the reduced (16 item) posttest would order the students in the same manner as the entire 16 item test, a cluster analysis was computed using the item scores for each student on both Form A and B of the 16 item instrument. The computer program used was the V-analysis portion of the BCTRY cluster analysis written by Robert C. Tryon, University of California, and Daniel E. Bailey, University of Colorado.⁸

The results for posttest Form B produced 2 clusters consisting of 4 items each (see Figure 1).

⁸For a discussion of cluster analysis, see Robert C. Tryon, "Cumulative Communnality Cluster Analysis", Educational and Psychological Measurement 18:3-35, 1968.



CLUSTER 1 = 6, 8, 14, 27 CLUSTER 2 = 22, 10, 9, 15

Figure 1. Cluster Analysis, Form B.

The first (most independent) cluster consisted of items 6, 8, 14, and 27, with item 6 as the pivot variable. The second cluster (Figure 1) consisted of items 22, 10, 9, and 15, with item 22 as the pivot variable. Table 41 summarizes the objective measured by each item in cluster 1 and the cognitive and diagnostic levels measured. Table 42 summarizes the objective measured by each item in cluster 2 and the cognitive and diagnostic levels measured.

TABLE 41

Summary of Item Objectives, Level,
and Gain for Cluster 1, Posttest B

Item #	Objective	Diagnostic Level	Gain	Cognitive Level
6 Pivot	9-Students will locate and describe external evidence of pulse.	Baseline	68-88	Knowledge (Identifying)
8	2-Students will associate food types with plant and animal sources.	Baseline	93-97	Knowledge (Classifying)
14	3-Students will distinguish differences in physical characteristics of food.	Baseline	95-94	Knowledge (Classifying)
27	6-Students will relate functions of teeth, tongue, and saliva in the mouth to characteristics of food.	Baseline	76-88	Knowledge (Identifying)

TABLE 42
Summary of Item Objectives, Level,
and Gain for Cluster 2, Posttest B

Item #	Objective	Diagnostic Level	Gain	Cognitive Level
10	5-Students will observe and describe the structure and function of mouth parts.	Baseline	72.5-92.0	Knowledge (Identifying)
22 Pivot	11-Students will associate heart actions with heart sounds and pulse.	Achievement	58.0-75.5	Knowledge (Identifying)
9	7-Students will describe the location and function of the esophagus.	Achievement	28.5-57.5	Knowledge (Identifying)
15	5-Students will observe and describe the structure and function of mouth parts.	Achievement	57.5-67.0	Knowledge (Identifying)

The domain validities of the cluster scores for cluster one and cluster two are .8557 and .7930, respectively. The correlation between the two oblique cluster domains is .1847, indicating a high degree of independence between the two clusters. The reliability of cluster score one is .7218, and the reliability of cluster score two is .6288.

Table 43 summarizes the correlations between the items in each cluster with the oblique cluster domains.

TABLE 43

Correlations of Items with the
Oblique Cluster Domains, Posttest A.

Cluster 1		Cluster 2	
Item No.	r	Item No.	r
6 pivot	.8828	9	.4475
8	.7042	10	.6487
14	.4492	15	.4136
27		22	.6195

The results of the cluster analysis for posttest A also yielded two clusters. However, when the items were plotted on the surface of a sphere (see Figure 2) it was apparent that there was only one large cluster with one item (#13) lying outside of the cluster. Therefore, item 13 was eliminated and one large cluster was used in the interpretation of the results. The cluster contains items 6, 1, 26, 3, 7, and 12, with item six as the pivot variable.

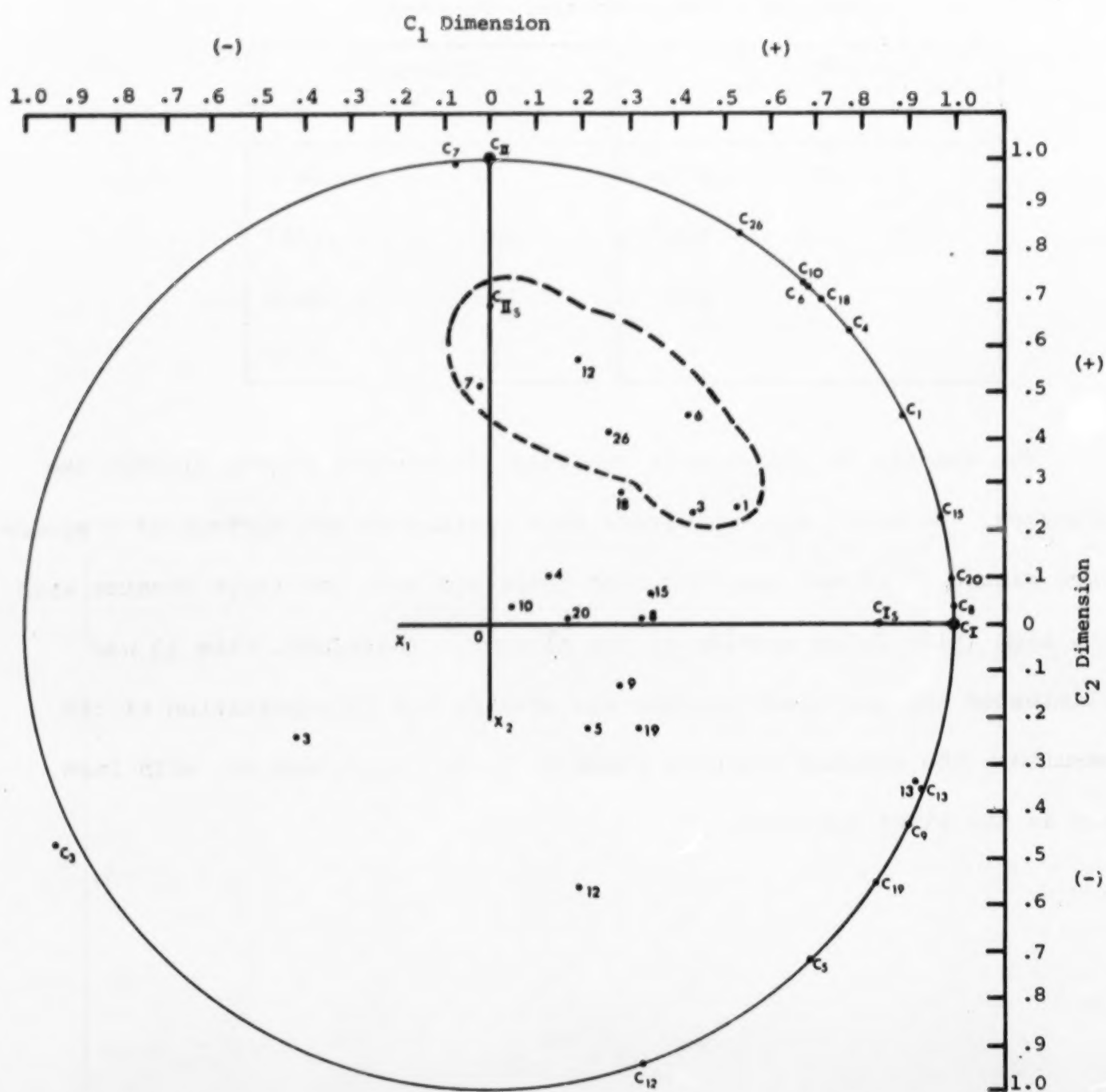


Figure 2. Cluster Analysis, Form A.

Table 44 summarizes the objective measured by each item in the cluster and the cognitive and diagnostic levels measured.

TABLE 44

Summary of Item Objectives, Level,
and Gain for Cluster 1, Posttest A

Item #	Objective	Diagnostic Level	Gain	Cognitive Level
6 Pivot	1-Students will associate food with generalized body needs.	Baseline	22-44	Knowledge (Identifying)
26	6-Students will associate circulation with distribution of food.	Achievement	35-30	Knowledge (Identifying)
3	5-Students will observe and describe the structure and function of mouth parts.	Baseline	96-94	Knowledge (Identifying)
7	9-Students will locate and describe external evidence of circulation.	Baseline	63-85	Knowledge (Identifying)
12	6-Students will relate functions of teeth, tongue, and saliva in the mouth to characteristics of food.	Baseline	76-89	Knowledge (Identifying)
1	1-Students will identify water as a substance present in many foods.	Baseline	62-74	Knowledge (Identifying)

The domain validity of the cluster score for this cluster is .6934.

The reliability of the cluster score is .6352.

Table 45 summarizes the correlation between the items and the oblique cluster domain on Form A.

TABLE 45
Correlations of Items with the
Oblique Cluster Domain, Posttest A

Item No.	r
6	.6215
26	.5009
3	-.4282
7	.3865
12	-.3400
1	.5298

Discussion

All of the items in cluster 1, posttest B, measured baseline data at the cognitive level of knowledge. The pivot variable was an item measuring an objective in circulation, and the remaining three items measure objectives in digestion. A general categorical title for cluster 1, posttest B, would be "Baseline-Selection."

Only one of the items in cluster 2, posttest B, measured baseline data (item 10). All of the other items were used to assess achievement, based on the pretest-posttest gain. As in cluster 1, the pivot variable measured an objective in circulation and the other three items measured objectives in digestion. All were judged to function at the cognitive level of knowledge.

The major category being assessed by the four items in cluster 2 was function. The general categorical title for cluster 2, posttest B,

would be "Achievement-Function."

None of the items involving the higher cognitive processes were included in the two most independent cluster groups. Those items and clusters of items that ordered the students in the same manner as the reduced Posttest B 16-item subtest all measured the cognitive level of knowledge.

With the exception of item 26, all of the items in the posttest A cluster measured baseline data, and all measured the cognitive level of knowledge. Item 26 was judged to be an achievement item. The pivot variable, item 6, measured an objective concerning digestion, and all of the remaining items in the cluster, except item 7, measured digestion objectives. Item 7 measured a circulation objective. A general categorical title for the posttest A cluster would be "Baseline-Function and Identification."

Three item pairs from the two parallel test forms were included in the clusters. These item pairs were A-3, B-15; A-12, B-27; and A-7, B-6. Those item pairs that were not included in both forms were probably excluded because of nonparallel choices in the items.

With the exception of item 26, no item included in a cluster measured an objective beyond objective 11, and it is debatable whether or not item 26 was measuring objective 16 or objective 8. Since the most important part of the experimental materials was written for objectives 12-16, it is apparent that none of the items measuring these objectives functioned in ordering the students. In future testing, more emphasis should be placed on writing test items that measure objectives 12-16. Also, more emphasis should be placed on writing items that measure the

higher cognitive behaviors. This is especially important since one of the underlying assumptions with the BSCS E.M.H. materials is that the E.M.H. student is capable of performing at the higher cognitive levels.

Objective Achievement Tests: Discussion

The two forms of the objective achievement test are reasonably comparable as determined by reliability and content validity. The posttest reliabilities of 0.59 and 0.61 are low, but acceptable for 16-item instruments. A finding of great importance for further evaluation studies is knowledge that E.M.H. students can handle the item form of these instruments.

Students in the experimental group performed significantly higher on the posttest than the control group, indicating that the treatment effect, the curriculum, made a difference. The item data in Tables 24 to 40, however, show a pattern of control group gains on the correct responses of Form A, pretest to posttest, and a pattern of loss, or little gain, in the correct responses of Form B. This interpretation is based on a within test-forms, i.e., different student group, comparison. On Form A, the control group had a greater percentage of students selecting the correct response (or a lesser percentage of students moving away from the correct response) pretest to posttest, on nine questions, while the experimental group had a greater percentage of students selecting the correct response (or a lesser percentage of students moving away from the correct response) pretest to posttest, on seven items. On Form B the same comparison favors the experimental group on 15 items and the control group on two items.

In order to reduce the differential between forms, the pooled

comparison results in the experimental group exceeding the control group on nine item pairs and one item, and the control group on one item pair and one item. If the comparison is made across item forms (A pretest and B posttest, or B pretest, A posttest), i.e., the same student group, comparisons can be made on pretest to posttest changes from one item of an item pair to its analogue. The combined Form A and B results of this analysis yield a greater percentage of experimental group students selecting the correct response, pretest to posttest, on 20 items, with the control group exceeding the experimental group on 10 items.

The apparent anomaly in control group gain is also clarified by noting (see Table 8) that two classes (teachers 60 and 61) within the control group taking Form A as a posttest had large and atypical gains. These teachers indicated that they both taught human body internal structures in the period between pretesting and posttesting. If these two classes were removed in the item analysis, the control group would show little item gain. The subjects taught by other control group teachers isn't known.

The test contains a disproportionate number of items that serve a diagnostic function. Few items were found to assess achievement, but experimental classes did gain on these items. Results of the cluster analysis indicated that those items that ordered the students were all at the cognitive level of knowledge. Several important objectives were missed completely.

In summary, the evaluation of the revised materials should include objective measures using the format and style of Forms A and B, but with groups of items to assess achievement of each objective. Student success

on items such as A-9, B-17, and A-15, B-25, provide sufficient evidence of conceptual thinking. Students should have an increased opportunity to behave conceptually on the revised instruments.

A test-retest reliability of 0.50 was obtained from the pooled 32 item pretest-posttest scores of the control group. This figure is reasonably high considering that at least two of the control group teachers taught digestion and circulation between pretest and posttest administrations.

CHAPTER III

SUMMARY AND RECOMMENDATIONS

The major purpose of this formative evaluation is to provide data and interpretations that will enable the curriculum design staff and consultants to revise the experimental materials. This evaluation may also be considered as a test of the experimental design and instrumentation. Recommendations for design modifications are important outcomes of the study. Some interpretations have potential implications for challenging existing beliefs about E.M.H. youngsters and what they can accomplish in school.

The first part of this summary will present interpretations of the data relevant to each of the sixteen terminal objectives specified in Me Now. The second part will reconsider the questions proposed in Table 1⁹ presenting interpretations based on the data reported herein. The last part of the summary will present recommendations for curriculum materials revision, research design and instrumentation, and implementation.

Were the Objectives of the Program Achieved?

Objective 1. Students will associate food with generalized body needs.

The writers of the curriculum materials for Me Now, Unit I, assumed that the ideas to be developed in the early sections would be a review for intermediate grade (age 11 to 13) youngsters. Responses secured during the first classroom trials of the materials, as indicated by

⁹ See pages 5-8.

teacher reports, indicate that the ideas associated with this objective were easy for the students to discuss.

The operations of measuring height and weight were inefficient, as described in Me Now. The plotting of height and weight data on the graphs was not satisfactorily done by the students. Teacher intervention seemed to be required more when the height and weight activities were conducted simultaneously, than when separated. Either more time and thought (by students) need to be provided or specific steps for measuring and graphing should be developed. For example, a teacher strategy for enabling students to measure and plot height (or weight) on successive days, with a thorough discussion of what is being done and with instruction to enable students to think with the activity, would be one approach to the problem.

Objective 2. Students will associate food types with plant and animal sources.

Several teachers reported that students had difficulty responding with the term "plant" as a category for grouping food sources. They initially responded with categories such as fruit or vegetables. This difficulty was not reflected in the responses to item pair A-8, B-8.

In general the strategies and materials were effective (although several revisions were suggested), but achievement data were lacking on the objective. Teacher reports and the one item pair support the suggestion that this objective may need to be only brief reviewing; students have the concept, but the relationship of physical characteristics to chewing, digestion, health, etc., is entirely missing.

Objectives 3 and 4. Students will distinguish differences in physical characteristics of foods.
Students will identify water as a substance present in many foods.

The strategies related to these objectives were the first to be rated by over half of the teachers as only partly successful. The reason most highly selected was, "ideas too complex."

Activities 9 and 10 seem to bear only an oblique relationship to the objectives, since the central concern in the activities is with hard-soft and solid-liquid. The difficulty of interpreting the activity is clearly expressed in one teacher's comments. "...they had a hard time guessing that hard and soft foods depend on the amount of water." The material apparently attempts to lead to the development of these relationships: hard foods are dry foods are solid foods; soft foods are moist foods are liquid foods. These relationships are false, since some hard foods (cabbage) may be more solid, but contain more liquid than some soft foods (bananas). One problem is with subobjective 4-1, "students will associate the presence of water with softness and the liquid condition of foods."¹⁰ This oversimplification, with the inference of a single causal relation resulted in confusion. The objectives and materials need revision to ameliorate this difficulty.

Objective 5. Students will observe and describe the structure and function of mouth parts.

Teachers reported general success with the activities and strategies designed to teach this objective. The major qualification to reports of success was that the program was too easy for the students, resulting in a loss of attention. Alternative activities or patterns of instruction are recommended for areas of study in which students have prior knowledge.

The objective test data do not confirm the teacher reports of prior

¹⁰ Me Now, page 16.

knowledge, but do show substantial gains from pretest to posttest scores. Limitations of the test items to low cognitive levels illuminated the students' specific factual knowledge but did not yield information on more general knowledge of structure and function of mouth parts.

Objective 5 could be subsumed under Objective 6 with the activities suggested as optional. Diagnostic procedures would need to be provided for the teacher should this suggestion be implemented. The average time of over one hour devoted to this objective does not seem warranted in view of the available data.

Objective 6. Students will relate functions of teeth, tongue, and saliva in the mouth to characteristics of foods.

Teachers reported high student interest and success for this objective. Students participated actively in the activities specified in the materials. There was only one report that the materials were too easy; a judgment that is supported by the one pair of items designed to assess this objective.

This observation led the investigators to speculate about factors affecting teacher judgment of level of difficulty of learning activities. For example, students may appear to like an activity even though they already understand the objectives the activity was designed to develop. In this instance, teacher reports might overlook an activity that is too simple for the student group. Conversely, if students appear to dislike an activity, teachers may report inattention and comment that activity and objective are already known by the student group, a judgment that would need empirical verification.

Objective 7. Students will describe the location and function of the esophagus.

Objective 7 is much more restricted than most of the objectives prepared for this program. This limitation is at least partly reflected in the inability of the test writers to prepare two forms of one item, or more than one item. Two-thirds of the students did not have prior knowledge of this objective as assessed by items A-5, B-9.

Teacher reports reflected satisfaction with the strategies and in the interest shown by the students.

With this limited concept, high student interest, and a recall item on the object test, one might well anticipate optimum learning by students. However, success on item pair A-5, B-9 only increased from about 30 percent to 60 percent of the experimental group. Greater gains were made on item A-10, but the posttest achievement level was about the same for both items.

The question of the adequacy of learning is justifiably raised with this restricted objective.

Objective 8. Students will relate stomach functions to changes in food form.

The activities designed to develop the behavior specified for this objective were reported to be interesting and valuable to students.

Gains were made on the item pair designed to assess this objective, but additional items are needed to provide an effective assessment. Stomach function was assessed, but changes in food form were not.

The food tests were reported to have captured the interest and enthusiasm of the youngsters. However, a question was raised by one teacher about the development of the significance of the food tests to

the changes in food form demonstrated by the tests. Unfortunately, no objective data are available to determine the extent to which the significance of the food changes was attained.

This is the first activity set in which several teachers indicated that few or no students were able to perform the objective. The next round of field testing, after the materials have been revised, needs to include mechanisms for securing data on this problem.

Me Now does not seem to be clear about teacher demonstration, student demonstration, or student participation regarding the activities in this section. Practice varied among the teachers. The achievement of the objectives would seem to require active participation by each student if the multisensory learning assumption is to be implemented.

Objectives 9 and 10. Students will locate and describe external evidence of circulation.
Students will locate and describe evidence of heart activity.

Two kinds of difficulties were reported by teachers for the activities associated with these objectives. First, there was difficulty in hearing the heartbeat with the materials provided, and second, teachers could not be sure if students actually counted pulse beats, or faked a pulse count that they thought would be reasonable. Materials revision should take these problems into account.

Again, as in the activities of Objective 8, several teachers reported that few or none of their students could perform some of the behaviors specified.

Recognition of where pulse could be felt on the body was successfully

accomplished by nearly 70 percent of the students after instruction.¹¹

The critical causal relation between heartbeat and pulse beat was not assessed by the objective test, as the items designed to assess this relationship were rejected from the analysis.

Objectives 11 and 12. Students will associate heart actions with heart sounds and pulse.
Students will associate circulation of blood with the heart action and pulse.

Teachers reported favorably on the strategies and activities associated with these objectives. Reports of student success were also very positive.

Gains were not made on the one item pair relating to heart function. This may have resulted from the many lines on the drawings, as well as lack of student achievement.¹²

Teachers reported that students suggested that the heart pumped air in initial discussions of heart function. This observation is supported by a selection of air for what the heart pumps by one-third of the students on the pretest. The effectiveness of instruction may be questioned (even at the recall level), however, as nearly 20 percent of the students selected air on the posttest.

There were no useful items on the objective test to assess Objective 12.

¹¹See Table 5, page 18.

¹²See page 58.

Objectives 13 and 14. Students will construct and explain a simplified model of a circulatory system.
Students will infer a relationship between food and blood and describe barriers between them.

Teachers reported success with the attainment of Objective 13. Some difficulties were noted with the psychomotor aspects of the activities for this objective. No test items were prepared to assess this objective.

Objective 14 is a critical link in relating the movement of digested food to the blood. Four of eight teachers reported that only a few of their students suggested that digested food might go into the blood. Teacher reports suggested increased diversity in class responses to the activities. For some classes, the activities were interesting reviews, for others they were boring, and for others, they seemed almost like new ideas.

No objective measures were retained that bear directly on Objective 14. The one related item (B-13) showed gains, but posttest achievement was very low.¹³

At this point in the materials, writers should give special attention to alternative pathways for teaching strategies and pupil activities. The quality of prior learning and retention can not be assumed to be uniform either within or among class groups. At points of review and synthesis, strategies for immediate diagnosis and alternative pathways would seem to provide for the variations that seem to inevitably occur in the instructional process.

¹³ See page 40.

Objectives 15 and 16. Students will observe and describe movement of solutions through membrane barriers. Students will associate circulation with distribution of food.

Only eight of the eleven experimental materials teachers completed the materials by the end of the school year. Therefore, the number of reports (and also the extensiveness of teacher comments) were considerably reduced for these objectives.

Three teachers rated the strategies as only partly successful. Comments regarding student responses infer to the investigators that the number of students relating the membrane demonstrations to food in the body and in the blood was quite limited.

Responses to two item pairs (A-15, B-25 and A-26, B-26) confirm the inference above, in that average posttest achievement on the four application items was only 32 percent correct.¹⁴

What Were the Results of the Formative Evaluation?

This section of the summary will follow the sequence of questions presented in Table 1 that served as a guide to the formative evaluation. Questions will not be repeated, but will be answered as best they can in terms of the data presented in this report.

The first field trials of Me Now were conducted with the range of classes that met the criteria for the target population proposed by the curriculum designers. Both experimental and control group youngsters were able to perform many objectives of the instruction prior to the beginning of the formative evaluation as measured by the objective test.

¹⁴ See pages 63, 64.

For example, for ten item pairs on the two alternate pretest forms, more than 50 percent of the students were able to respond correctly. This finding was contradictory to the opinions expressed by many individuals concerned with special education, and by the opinions of the curriculum development staff prior to securing pretest data. The general assumption had been that these youngsters had not studied the internal parts of the body, that the concepts were too abstract for them to handle, and that they had very little of what could be called academic learning in their school experience.

It is apparent that the students did well on the large proportion of those items of the pretest that were retained for analysis. The pretest turned out to be gathering baseline data rather than providing opportunities to show gain and achievement.

Prescribed teacher strategies are an essential feature of the materials developed in this project. Data from teacher reports indicate that teachers followed the strategies in most cases, modified them in a more limited number of instances, and replaced them in an even more limited number of cases. These reports are contradicted by informal observations of video tapes made in classes of two of the experimental teachers. Formal analyses have not been conducted of these tapes, although a limited study is underway at the writing of this report. However, BSCS staff review of many tapes permits the interpretation that though the strategies were followed, the intent of the strategies was missed in many cases. This intent refers to the model of inquiry as conceptualized by the curriculum development staff. In this model the role of the teacher is to provide phenomena and experience that will require student

interpretation, analysis, inference, prediction and testing. This model implies a reduction in the amount of teacher talk, especially teacher telling, and an increase in active verbal participation by students. Provision must, therefore, be made for student responses, for replies, and for challenges to each other's responses and replies.

In many instances time was not provided for more than one student to respond to a particular problem. In many cases demonstrations were conducted rather than individual or small group student experimentation. The intents were apparently not clearly delineated in Me Now. In many instances teachers asked questions and did not give students time to think and then respond. Rather, they presented a continuous flow of verbal discourse, following question with question or repeats of the question, rather than providing the students with a period of silence after a question so that the question could be thought out and a response suggested. The materials provided no prompts to the teacher to suggest that after a student response is made other students might be called upon to comment on the initial response. In this way student interaction could be promoted. This intention was clearly in the minds of the curriculum development staff but does not appear anywhere in Me Now.

Informal reports of teachers and discussions with staff visitors to classes indicated that teachers felt pressured to push on through the materials in order to complete them by the end of the school year. Even with this push, only eight of the eleven teachers completed the materials. The greatest amount of time available to any experimental teacher was eight weeks. The estimate of four to six weeks, at thirty minutes per day, hypothesized by the development staff was clearly an underestimate.

A more reasonable estimate is that these materials, properly paced and timed to allow maximum student participation at each stage of development, require at least a full quarter of school work. In all probability, as teachers modify the materials as they gain experience, the program might well be taught over a period of a full semester. This would be especially important if the level of achievement is to be raised.

For the most part, teacher reports indicated that the intended student behaviors did occur. In several sections of the materials student interest lagged in some classes. No reports indicated a lack of interest across all classes for any one activity. It is apparent that the range of activities needs to be extended to adapt more fully to student differences within class groups.

Very little information was secured to determine that the behaviors occurred to an acceptable degree. With this initial testing the development staff was even uncertain if they had reached the appropriate level of the target population. The data are considered to support a general statement that, in the first half of the materials, student behavior did occur to an acceptable degree and perhaps the materials were oversimplified for the student population. In the last half of the materials, and especially, the last third, this problem is more difficult to appraise. First, there are not enough data, and second the definition of acceptable degree has not been formulated. A limited number of objective test items and teacher reports lead the investigators to propose that in the latter third of the materials student behavior did not occur to an "acceptable" degree. This may have been a result of the rapidity of instruction and the inopportunity for many students to respond at all to the behaviors.

In only a few instances were the materials considered too difficult for the maturity and psychophysical development of the test population. Again, from observation of video tapes, lack of attention seemed more common than that reported by teachers. However, inattention is difficult to judge inasmuch as many of these students appear to be inattentive, but when asked a question they were able to respond easily. A major limitation on student behaviors was the lack of clearly communicating to teachers that each student be given an opportunity to perform the behavior present in Me Now.

Precise relationships between teacher-behavior and student-behavior cannot be discussed because no data were collected.

The experimental group teachers were enthusiastic about both the materials provided to accompany Me Now and the strategies presented with the materials. Accounts of increased interest and enthusiasm on the part of students and reports of the effectiveness of particular materials were frequently reported throughout the evaluation period. In interview discussions with experimental teachers a feeling was expressed that teachers felt constrained to follow the prescribed strategies of the materials and that if they were to teach the materials again they would not feel as constrained and would make many more modifications than they did during this initial trial period.

Teachers' reports generally indicated that students were interested throughout the study. The major student reaction on termination of the study was disappointment that the class had completed this part of their study or that they would not be able to finish before the end of the school year. A few responses from parents also indicated that enthusiasm

carried over to the home and stimulated many discussions in the homes of some children regarding the functioning of the human body.

In final summary, the evaluators conclude, that students met many objectives of the program, but not to the extent desirable. The last phrase must be qualified because the program has provided enough evidence to suggest that "extent desirable" can now be given a new definition. The project sights can be raised and more E.M.H. students can be expected to gain more from the instructional program. The clear differences between the experimental and control groups (significant at the .005 level) on an objective measure, also being evaluated, hint at the potential for growth in this student population.

The statement, "not to the extent desirable", must also be qualified by the limitations in opportunities for students to perform objectives during the course of instruction. The curriculum development staff has not, in these materials, committed itself to any particular level of desired student achievement. The investigators urge that as the materials are revised and additional materials in the Life Sciences for the Educable Mentally Handicapped program is continued, the level of desired student success be specified.

The content and sequence of the materials seem satisfactory and in no need of revision. More activities involving higher thought processes can be included in the revised materials. Also, materials revision might well include increased attention to divergent questions and activities as against convergent questions, which predominate the existing materials.

A major finding of this formative evaluation is the development and successful use of an achievement test instrument in multiple choice format

that is independent of both measured intelligence and reading ability. Teacher reports indicated that students enjoyed the objective test and were, for the first time during testing, at ease because the format provided an easy means for them to keep their places within the test. The directions for student responses were simple so that they were more certain that they had marked each response as they intended it to be marked. This limited test also provides a hint that conceptual thinking is possible in this student population. Greater emphasis in future test development can be placed on higher cognitive processes and on assessing terminal rather than enabling objectives of the materials.

A major focus of the Biological Sciences Curriculum Study has been on the processes of inquiry and their importance as a necessary element of science teaching. No data were gathered regarding these processes as a part of this evaluation study. The development staff has a responsibility to be more explicit regarding their intentions in terms of both student and teacher behaviors and the inquiry aspects of instruction. The clarification of teacher behaviors that support or undermine the processes of inquiry is a necessary step in the revision of these materials and any further materials development. Videotaping, even for an informal review, has been shown to make a valuable contribution to provide a curriculum development group with data that is necessary to focus their attention on the degree to which intentions have been communicated to teachers, through the materials, or through the teacher briefing sessions.

Implementation of this curriculum apparently requires a rather drastic change in teacher behavior. Special education teachers, as

viewed by the investigators, tend to work with individuals and small groups. These materials require a teacher-total class involvement that is unusual for these teachers. The inquiry mode itself, the value of utilizing situations to allow students to make sense out of their immediate environment, seemed to be foreign to the experimental group teachers. There seems to be a common operational principle to tell students answers to their questions in this teacher group. Observation protocols and observers in the classroom should be a part of the next round of evaluation studies with materials in this program. Observers, using protocols, will facilitate data reduction and analysis, whereas video tape analysis requires extended and expensive monitoring time after the video tapes have been secured.

Appropriate teacher behavior is crucial to achieve the intentions of this curriculum. Evaluation studies of the revised materials should include a search for questions regarding problems of implementation. Without proper implementation the materials as currently designed can easily be warped into a lecture discussion model, a model that has failed with so many youngsters who are more verbally able than the target population.

What is Recommended?

Materials Revision

1. Implementation of the inquiry model should be made more explicit throughout the materials. The questioning strategy needs to be made explicit. Examples of behaviors that support and undermine inquiry may be helpful.

2. Frequent review activities that expand and apply the desired behaviors should be incorporated into the materials. Attention should be given to preparing questions and activities to levels above recall, e.g., interpretation, application, etc.
3. Several early sections of Me Now should be upgraded or extended to provide for students with high entry behaviors.
4. Suggestions that teachers give students time to think, to wait after asking a question, and to encourage student-student reaction should be incorporated into the materials.
5. Modifications of some of the materials provided for both students and teachers should be revised, incorporating specific suggestions of the experimental group teachers.
6. No general information or overviews are provided the student to assist him in understanding what the objectives are and why he is engaged in the study. Advance organizers (Ausubel) would seem appropriate for the teaching strategies. The day to day activities may seem meaningless to the students if they don't have a broad general structure into which the activities fit.
7. The statements of objectives were not directly subjected to analysis in this study. However, revision of several objectives has been suggested in the body of the report. Objectives should be examined for comparable hierarchial levels, i.e., the terminal objectives should all be conceptual and at the same level of generality and abstraction.
8. Attention should be given to modes of teacher preparation for using this curriculum that supplement Me Now. These might

include training films, simulation exercises, or training in the use of the classroom observation protocols to be developed. They also could entail a more briefly structured orientation/training workshop prior to using the curriculum or short training sequences provided during the use of the materials.

Achievement Testing

1. The feasibility and value of achievement testing has been demonstrated in this report. Achievement items at higher cognitive levels than recall should be constructed. Conceptual complexity is not directly related to difficulty. Items that can be understood, but are conceptually more complex, are the desired goal.
2. The achievement tests used in the current evaluation lacked effective items to assess the most important objectives of the instructional program. Priority should be given to the development of items for Objectives 12 to 16.
3. Performance tests that are not dependent on paper and pencil should be developed.
4. Measures for assessing achievement of inquiry behaviors must be developed to evaluate the efficacy of the inquiry model for the E.M.H. population.

Evaluation Design

1. Background data regarding father's occupation, and evidence of cerebral dysfunction can be eliminated.

2. Teacher reports of student and home conditions that may influence learning should be secured.
3. No data were gathered on teacher characteristics in the 1970 evaluation. Teacher input is an important variable in the instructional program. Administration of instruments such as the Meyers-Briggs scale, or Harvey's teacher characteristics, should be included in the next evaluation program.
4. Observation protocols that will make it possible to secure data on the questioning behaviors of teachers and response patterns of students should be developed.
5. Observation protocols should provide for recording unplanned student and teacher behaviors.
6. The feasibility of interviewing students and the preparation of interview schedules should be considered in the next evaluative study.
7. The many issues raised in the 1970 field trials suggest to the evaluators that intensive, rather than extensive, evaluation be continued with this curriculum project. Different methods of gathering overlapping data is strongly encouraged. For example, observers, audio and videotaping might be used in series and simultaneously.
8. Implications of student-teacher transactions for teacher preparation should be an important aspect of data analysis.
9. A questionnaire should be designed to secure information from control group teachers regarding their instruction that could influence behaviors specified in the experimental program.

REFERENCES CITED

- Baker, Frank B., and Martin, J. J. Fortap: A Fortran Test Analysis Package, Laboratory of Experimental Design. Wisconsin Research and Development Center for Cognitive Learning. The University of Wisconsin, Madison, Wisconsin, March 1, 1968.
- Biological Sciences Curriculum Study. Me Now, Unit 1, Digestion and Circulation (Experimental Edition). Boulder, Colorado: The Regents of the University of Colorado, Spring, 1970.
- Hopkins, Kenneth D. "Regression and the Matching Fallacy in Quasi-Experimental Research." Mimeographed. Boulder, Colorado: Laboratory of Educational Research, University of Colorado, n.d.
- Stake, Robert E. "The Countenance of Educational Evaluation." Teachers College Record 68:523-540 (April), 1967.
- Tryon, Robert C. "Cumulative Communality Cluster Analysis." Educational and Psychological Measurement 18:3-35, 1968.